

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:02:42 ; Search time 2891.94 Seconds  
(without alignment) 10100.295 Million cell updates/sec

Title: US-09-835-992a-19

Perfect score: 714  
Sequence: 1 cgcagagaaaagtcatttca.....cagactgtctgcacattca 714Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pal:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_ey:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sce:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_hcg\_hum:\*  
39: em\_hcg\_mus:\*  
40: em\_hcg\_mus:\*  
41: em\_hcg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710	99.4	714	6	AR146579
2	710	99.4	714	6	BD079828
3	672.8	94.2	1052	9	HSSCPX16
4	672.8	94.2	2572	2	HUMSCP2A
5	672.8	94.2	12176	2	AL358233
6	672.8	94.2	175044	2	AC027278
7	672.8	94.2	193774	9	AL445183
8	667.6	93.5	1439	9	BC005911
9	611.8	85.7	687	6	AR146580
10	611.8	85.7	687	6	BD079829
11	474.6	66.5	1229	9	HS2450
12	459.8	64.4	1219	9	HUMSTREA
13	436.2	61.1	843	6	AR146582
14	436.2	61.1	843	6	BD079831
15	318.4	44.6	1500	9	HUMSCP2B
16	250.6	35.1	2661	4	AF051897
17	202.8	28.4	2599	10	RAT60KDA
18	194.6	27.3	14554	2	AC134794
19	194.4	27.2	2571	10	RATSCPXA
20	193.2	27.1	2626	10	BC018384
21	193.2	27.1	248677	2	AL844206
22	169.8	23.8	904	10	S80339
23	165.6	23.2	173	6	AR246194
24	107.4	15.0	263	6	A74403
25	107.4	15.0	263	6	A77382
26	89.2	12.5	151259	2	AC138822
27	89.2	12.5	160701	9	AC108106
28	89.2	12.5	172404	9	AC138948
29	89.2	12.5	181154	2	AC138842
30	89.2	12.5	185785	9	AC138817
31	89.2	12.5	188885	9	AC138850
32	89.2	12.5	192925	2	AC138953
33	89.2	12.5	201419	9	AC138908
34	89.2	12.5	211297	2	AC138860
35	89.2	12.5	216200	2	AC138862
36	89.2	12.5	225384	2	AC138832
37	89.2	12.5	238249	2	AC138960
38	87.6	12.3	130416	9	AC008851
39	85.2	11.9	113202	9	HSJ397H23
40	79.8	11.2	149597	2	AC034271
41	79.8	11.2	157749	2	AC025535
42	78.8	11.0	273275	3	AE014828
43	77.8	10.9	8056	6	AX599046
44	77.4	10.8	125623	3	AC115599
45	77.2	10.8	37957	8	U17009

## ALIGNMENTS

RESULT 1  
LOCUS AR146579 714 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 19 from patent US 6218521.  
ACCESSION AR146579  
VERSION AR146579.1 GI:15109768  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 714)  
AUTHORS Obata,Y.  
TITLE Isolated nucleic acid molecules associated with gastric cancer and methods for diagnosing and treating gastric cancer  
JOURNAL Patent: US 6218521-A 19 17-APR-2001.

FEATURES Location/Qualifiers  
 source 1..714  
 /organism="unknown"  
 BASE COUNT 258 a 100 c 92 g 260 t 4 others  
 ORIGIN

Query Match 99.4%; Score 710; DB 6; Length 714;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-93;  
 Matches 714; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CGCCGAAAAAGTTAATTTAATTTCTATTAACATCTCTCAAGCATTAATTTATCC 60  
 1 CGCCGAAAAAGTTAATTTAATTTCTATTAACATCTCTCAAGCATTAATTTATCC 60  
 61 TATATCTCACTGAATTTAAGAAATACATCTGTATTTGAAAACTAGAAAAAGATA 120  
 61 TATATCTCACTGAATTTAAGAAATACATCTGTATTTGAAAACTAGAAAAAGATA 120  
 121 ATGCAGATTAATTAACCTTACATGAAAAAGAAATTTATACAAAGAGCTGAGACCTTAT 180  
 121 ATGCAGATTAATTAACCTTACATGAAAAAGAAATTTATACAAAGAGCTGAGACCTTAT 180  
 181 AAATGAAATGAGATTAATTAATTTGAAACTGCATCTGAAAGCAACTTTATTTCAAT 240  
 181 AAATGAAATGAGATTAATTAATTTGAAACTGCATCTGAAAGCAACTTTATTTCAAT 240  
 241 ATNCTTAATGAGTGCTTTATGACTAATACCTGATTTTGAAGAAGAAACCATGTT 300  
 241 ATNCTTAATGAGTGCTTTATGACTAATACCTGATTTTGAAGAAGAAACCATGTT 300  
 301 AAAAATATTTTATTTAATAAATTAAGCCTGTCTGATCATATTTCTTTAT 360  
 301 AAAAATATTTTATTTAATAAATTAAGCCTGTCTGATCATATTTCTTTAT 360  
 361 TTGATTTGGGAAANAATCTGTTCTGATACATGAAATGCAAAATTTTATGATTTTA 420  
 361 TTGATTTGGGAAANAATCTGTTCTGATACATGAAATGCAAAATTTTATGATTTTA 420  
 421 ATCTCACTAATTTAANAATCTATGAAATGATTAATGACATGAGCAACACATA 480  
 421 ATCTCACTAATTTAANAATCTATGAAATGATTAATGACATGAGCAACACATA 480  
 481 ATTTACTGCGCAGCTGTGGCATTTGTTTCTTACTTACTTACTTACTTACTTACTT 540  
 481 ATTTACTGCGCAGCTGTGGCATTTGTTTCTTACTTACTTACTTACTTACTTACTT 540  
 541 AATTGAATCTTCAGAGAAATATCTTAAATATCTTTGTAAGCAAAACAAAGCTTTT 600  
 541 AATTGAATCTTCAGAGAAATATCTTAAATATCTTTGTAAGCAAAACAAAGCTTTT 600  
 601 TGTTTACATGATCTTTGGGATTTTACGTTCTTAATTTTATTTGAAACCTCAATTTTAC 660  
 601 TGTTTACATGATCTTTGGGATTTTACGTTCTTAATTTTATTTGAAACCTCAATTTTAC 660  
 661 CCCAGACCATTAATTAACATTAATTTGTTTGGACAGTTGTTGCAATTTCA 714  
 661 CCCAGACCATTAATTAACATTAATTTGTTTGGACAGTTGTTGCAATTTCA 714  
 661 CCCAGACCATTAATTAACATTAATTTGTTTGGACAGTTGTTGCAATTTCA 714

RESULT 2  
 LOCUS BD079828 714 bp DNA linear PAT 27-AUG-2002  
 DEFINITION Cancer-associated nucleic acids and polypeptides.  
 ACCESSION BD079828  
 VERSION BD079828.1 GI:22625431  
 KEYWORDS JP 2001516009-A/494.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 714)  
 Old, L. J., Scanlan, M. J., Stockert, B., Gure, Y. T., Gout, I.,  
 Ogbare, M., Obata, Y., Pfeundschnuh, M., Tureci, O. and Sahin, U.

TITLE  
 JOURNAL  
 LUDWIG INSTITUTE FOR CANCER RESEARCH  
 COMMENT  
 OS Homo sapiens (human)  
 PN JP 2001516009-A/494  
 PD 25-SEP-2001  
 PF 15-JUL-1998 JP 2000503425  
 PR 17-JUL-1997 US 08/896164, 10-OCT-1997 US 60/061599 PR  
 10-OCT-1997 US 60/061765, 10-OCT-1997 US 08/948705 PR  
 11-OCT-1997 GB 9721697.2, 22-JUN-1998 US 09/102322 PI  
 J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI  
 CHEN, IVAN GOUT, MICHAEL O'HARE, YUICHI OBATA, MICHAEL PFEUNDSCHUH, PI  
 OZLEM TURECI,  
 PI UGUR SAHIN  
 PC G01N33/574, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61K48/00, PC  
 A61P35/00,  
 PC C07K14/82, C07K16/32, C12N15/09//C07K16/46, C12P21/08, A61K37/02,  
 PC C12N15/00  
 CC Cancer-associated nucleic acids and polypeptides. FH Key  
 Location/Qualifiers  
 FT source 1..714  
 FT /organism="Homo sapiens (human)".  
 FEATURES Location/Qualifiers  
 source 1..714  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 258 a 100 c 92 g 260 t 4 others  
 ORIGIN

Query Match 99.4%; Score 710; DB 6; Length 714;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-93;  
 Matches 714; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CGCCGAAAAAGTTAATTTAATTTCTATTAACATCTCTCAAGCATTAATTTATCC 60  
 1 CGCCGAAAAAGTTAATTTAATTTCTATTAACATCTCTCAAGCATTAATTTATCC 60  
 61 TATATCTCACTGAATTTAAGAAATACATCTGTATTTGAAAACTAGAAAAAGATA 120  
 61 TATATCTCACTGAATTTAAGAAATACATCTGTATTTGAAAACTAGAAAAAGATA 120  
 121 ATGCAGATTAATTAACCTTACATGAAAAAGAAATTTATACAAAGAGCTGAGACCTTAT 180  
 121 ATGCAGATTAATTAACCTTACATGAAAAAGAAATTTATACAAAGAGCTGAGACCTTAT 180  
 181 AAATGAAATGAGATTAATTAATTTGAAACTGCATCTGAAAGCAACTTTATTTCAAT 240  
 181 AAATGAAATGAGATTAATTAATTTGAAACTGCATCTGAAAGCAACTTTATTTCAAT 240  
 241 ATNCTTAATGAGTGCTTTATGACTAATACCTGATTTTGAAGAAGAAACCATGTT 300  
 241 ATNCTTAATGAGTGCTTTATGACTAATACCTGATTTTGAAGAAGAAACCATGTT 300  
 301 AAAAATATTTTATTTAATAAATTAAGCCTGTCTGATCATATTTCTTTAT 360  
 301 AAAAATATTTTATTTAATAAATTAAGCCTGTCTGATCATATTTCTTTAT 360  
 361 TTGATTTGGGAAANAATCTGTTCTGATACATGAAATGCAAAATTTTATGATTTTA 420  
 361 TTGATTTGGGAAANAATCTGTTCTGATACATGAAATGCAAAATTTTATGATTTTA 420  
 421 ATCTCACTAATTTAANAATCTATGAAATGATTAATGACATGAGCAACACATA 480  
 421 ATCTCACTAATTTAANAATCTATGAAATGATTAATGACATGAGCAACACATA 480  
 481 ATTTACTGCGCAGCTGTGGCATTTGTTTCTTACTTACTTACTTACTTACTTACTT 540  
 481 ATTTACTGCGCAGCTGTGGCATTTGTTTCTTACTTACTTACTTACTTACTTACTT 540  
 541 AATTGAATCTTCAGAGAAATATCTTAAATATCTTTGTAAGCAAAACAAAGCTTTT 600



Db	541	AATTGAATCTTCACGAGAAATAATCCTTAATATACTTTTAAAGCAAAACAAGTTTTT	600
Oy	601	TGTTTACATAGTCTTTGGGATTTTACGTGTCCTAATTTTATTCGAAATCAATTTTAC	660
Db	601	TGTTTACATAGTCTTTGGGATTTTACGTGTCCTAATTTTATTCGAAATCAATTTTAC	660
Oy	661	CCCGAGACCATTAAATGCATATTTACTTTGTNTGACAGTGTGTTGCCAATTCA	714
Db	661	CCCGAGACCATTAAATGCATATTTACTTTGTNTGACAGTGTGTTGCCAATTCA	714

RESULT 3  
HSCSPTX16/c

LOCUS	HSCSPTX16	1052 bp	DNA	linear	PRI 25-JAN-1997
DEFINITION	Human sterol carrier protein-X/sterol carrier protein-2				
ACCESSION	(SCP-X/SCP-2) gene, exon 16, and complete cds.				
KEYWORDS	U11313.1	GI:532077			
SOURCE	16 of 16				
SEGMENT	Homo sapiens (human)				
ORGANISM	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 1052) Ohta,T., Renneft,H., Pfeiffer,S.M., He,Z., Yamamoto,R., Holt,J.A., Billheimer,J.T. and Straus,J.F. III. The structure of the human sterol carrier protein X/sterol carrier protein 2 gene (SCP2)				
TITLE	Genomics 24 (2), 370-374 (1994)				
JOURNAL	MEDLINE 95213031				
PUBLISHED	2 (bases 1 to 1052) Straus,J.F. III. Direct Submission Submitted (24-JUN-1994) Jerome F. Straus III, Department of Obstetrics and Gynecology, Division of Reproductive Biology, University of Pennsylvania Medical Center, 778 Clinical Research Building, 422 Curie Boulevard, Philadelphia, PA 19104-6142, USA				
FEATURES	Location/Qualifiers				
SOURCE	1..1052				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/map="lp32"				
	/tissue_type="liver"				
gene	join(U11297.1:1..3666,U11299.1:1..108,U11300.1:1..122, U11301.1:1..1182,U11302.1:1..115,U11303.1:1..177, U11304.1:1..114,U11305.1:1..137,U11306.1:1..201, U11307.1:1..198,U11308.1:1..158,U11309.1:1..1725, U11310.1:1..153,U11311.1:1..180,U11312.1:1..1301..1034 /partial				
mRNA	/gene="SCP-X/SCP-2" join(U11297.1:3452..3641,U11299.1:1:26..83,U11300.1:1:26..97, U11301.1:1:26..157,U11302.1:1:26..90,U11303.1:1:26..152, U11304.1:1:26..89,U11305.1:1:26..112,U11306.1:1:26..176, U11307.1:1:26..173,U11308.1:1:26..133,U11309.1:1:1547..1700, U11310.1:1:26..128,U11311.1:1:26..155,U11312.1:1:26..105, 26..1034)				
CDS	/gene="SCP-X/SCP-2" /product="sterol carrier protein-X/sterol carrier protein-2" join(U11297.1:3573..3641,U11299.1:1:26..83,U11300.1:1:26..97, U11301.1:1:26..157,U11302.1:1:26..90,U11303.1:1:26..152, U11304.1:1:26..89,U11305.1:1:26..112,U11306.1:1:26..176, U11307.1:1:26..173,U11308.1:1:26..133,U11309.1:1:1547..1700, U11310.1:1:26..128,U11311.1:1:26..155,U11312.1:1:26..105, 26..121) /gene="SCP-X/SCP-2" /codon_start=1 /product="sterol carrier protein-X/sterol carrier protein-2"				

Query Match	Best Local Similarity	Score	DB	Length
Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2	94.2%; 98.2%;	672.8; 3e-88;		1052;
<p>protein-2"</p> <p>/protein_id="AAB41286.1"</p> <p>/db_xref="GI:1773239"</p> <p>/translation="MSSSPWEPALRRVFFVGMGTFKFKPGAENSRDYPDLAEAGIKALDAIPYSAVDACVGVYGDSTGCGARIVHSLGMTGIPINUNNNCATGTSPALMARLIGGVAECVLTALGPEKMSKGIKISDSPTIPDHKVDLLINVGYSAPHVALOMEPYAKREHEKVGKTLIEHFAKIGKMGKHNHNVNPPVSGPDESDVMSKEYFDLTLIQCPSTDGAATAILASFAFVKQGLGSOAKAYELAQEMTDLPSSFEKSIITKIDFTGSKMARCYEKSGULPNDIDIVIELHDFSTNELITYEALGCEPGQATVLVDRNENIGKMAIRVINSGLISKIHPLATGIAQCAELCQWLRGAGRGROYGAVAAUADRDIAGAVVVTLYKMGFPEASAFRTHOIEAVNTSSADPKFNLVFEKTEKLEEGEGEVFKVIGIPAFKVPKVDGPGKEATVVDVNGKSVLPNSDKKACCTITMADSDFLALMTGRKNPSAFPGKLTKTGMNGLMKLQNLQDPGNKL"</p> <p>26.1034</p> <p>/gene="SCP-X/SCP-2"</p> <p>/note="nucleotides 122-1027 are not translated"</p> <p>/number=16</p>				
BASE COUNT	352 a	157 c	164 g	379 t
ORIGIN				
Query Match	Best Local Similarity	Score	DB	Length
Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2	94.2%; 98.2%;	672.8; 3e-88;		1052;
<p>protein-2"</p> <p>/protein_id="AAB41286.1"</p> <p>/db_xref="GI:1773239"</p> <p>/translation="MSSSPWEPALRRVFFVGMGTFKFKPGAENSRDYPDLAEAGIKALDAIPYSAVDACVGVYGDSTGCGARIVHSLGMTGIPINUNNNCATGTSPALMARLIGGVAECVLTALGPEKMSKGIKISDSPTIPDHKVDLLINVGYSAPHVALOMEPYAKREHEKVGKTLIEHFAKIGKMGKHNHNVNPPVSGPDESDVMSKEYFDLTLIQCPSTDGAATAILASFAFVKQGLGSOAKAYELAQEMTDLPSSFEKSIITKIDFTGSKMARCYEKSGULPNDIDIVIELHDFSTNELITYEALGCEPGQATVLVDRNENIGKMAIRVINSGLISKIHPLATGIAQCAELCQWLRGAGRGROYGAVAAUADRDIAGAVVVTLYKMGFPEASAFRTHOIEAVNTSSADPKFNLVFEKTEKLEEGEGEVFKVIGIPAFKVPKVDGPGKEATVVDVNGKSVLPNSDKKACCTITMADSDFLALMTGRKNPSAFPGKLTKTGMNGLMKLQNLQDPGNKL"</p> <p>26.1034</p> <p>/gene="SCP-X/SCP-2"</p> <p>/note="nucleotides 122-1027 are not translated"</p> <p>/number=16</p>				
BASE COUNT	352 a	157 c	164 g	379 t
ORIGIN				

RESULT 4  
HUMSCP2A/C 2572 bp mRNA linear PRI 06-DEC-1993  
DEFINITION Human sterol carrier protein X/sterol carrier protein 2 mRNA,  
complete cds.  
ACCESSION M75883  
VERSION M75883.1 GI:432974  
KEYWORDS sterol carrier protein-2, sterol carrier protein X.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 2572)  
AUTHORS George, H., Billheimer, J.T., and Straus, J.F. III.  
CDNAs encoding members of a family of proteins related to human  
sterol carrier protein 2 and assignment of the gene to human  
chromosome 1 p21---pter  
JOURNAL DNA Cell Biol. 10 (6), 559-569 (1991)  
MEDLINE 92029618  
PUBMED 1718316  
REFERENCE 2 (bases 1 to 2572)  
AUTHORS Vesa, J., Hellsten, E., Branoski, B.L., Emanuel, B.S., Billheimer, J.T.,  
Mead, S., Cowell, J.K., Straus, J.F. III, and Peltonen, L.  
Assignment of sterol carrier protein X/sterol carrier protein 2 to  
1p32 and exclusion as the causative gene for infantile neuronal  
ceroid lipofusiosis  
JOURNAL Unpublished  
COMMENT On Dec 6, 1993 this sequence version replaced gi:410029.  
FEATURES  
source Original  
location/Qualifiers  
1..2572  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="1p32"  
/tissue\_type="liver"  
/clone\_lib="Clontech HL-1001b;HL-1115b"  
1..2572  
/gene="SCP-X/SCP-2"  
22..1665  
/gene="SCP-X/SCP-2"  
/codon\_start=1  
/product="sterol carrier protein X/sterol carrier  
protein-2"  
protein\_id="AAA03557.1"  
/db\_xref="GI:432975"  
translation="MSSSPWEPATLRVGVVGVMTKEFKPGAEASRDYDPLAEAGK  
KALADAOIPIYSAVDACVGVYRGDSTCGORAIYHSLGMPPIINNNNCAIGSTALF  
MAROLIIGGVAVACVAILGPEKMSKSLIGKPSDRITLPDDKVDLLINVGSAHPVAP  
CMFGYAGKEHNEKTYGTIKIEHPAKIKRKHKSNNPYPQODEYSLDEVMASKEVPD  
LTLTCCPTSDGAATAIILASEAFVQKYGQSAVEILLQEMMTDLPSSFEESLITKMV  
GDMSEKARCKYERKGLTPNDIDVIELDCSTNELLYEALGLCPGQGAITLVDRG  
DNTYGGKAVINPSGGLISKGPLAGTACQCAELCMWREGKGVQAGAKVALDHNL  
GIGAVVTVLYMGPEPAASFRTHQIEVFPSSADGPKALVKELEKLEEGEO  
FYKTCGIFAFVKDGPGRKATWVDVKNKGSVLPSPDKKADCTITMADSPFLAM  
TGKMPQSAFPGSKLITGNMGLAKLQNLQPGNAL"

BASE COUNT 794 a 439 c 570 g 769 t  
ORIGIN  
Query Match 94.2%; Score 672.8; DB 9; Length 2572;  
Best Local Similarity 98.2%; Pred. No. 2.2e-88;  
Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2;  
/gene="SCP-X/SCP-2"  
/evidence=experimental  
2 GCCAGAAAAGTATTTTAAATTTTCTATTAACATCTCTCCAAAGCATTATTTATCT 61  
Db 2571 GCCAGAAAAGTATTTTAAATTTTCTATTAACATCTCTCCAAAGCATTATTTATCT 2512  
QY 62 ATATCTCCTGAATTTTAAAGAAATTAACATCTGTGTTGAAGAAATGAGAAATGATAA 121  
Db 2511 ATATCTCCTGAATTTTAAAGAAATTAACATCTGTGTTGAAGAAATGAGAAATGATAA 2452  
QY 122 TGCAGATTAATTAACCTTACATGAAAAGAAATTAATTAACAAAGACTGAGAACGTTATA 181  
Db 2451 TGCAGATTAATTAACCTTACATGAAAAGAAATTAATTAACAAAGACTGAGAACGTTATA 2392  
QY 182 AATTGAATGAGATTATTAATTTGAAAACCTGATCTGAAAGCAAACCTTATTTGTCATTA 241  
Db 2391 AATTGAATGAGATTATTAATTTGAAAACCTGATCTGAAAGCAAACCTTATTTGTCATTA 2332  
QY 242 TNCCTTAATGATGAGTTTATTAATGACCTAATACAGCTGATTTTCAAGAGAAACCATGTTA 301  
Db 2331 TNCCTTAATGATGAGTTTATTAATGACCTAATACAGCTGATTTTCAATAGAAACCATGTTA 2272  
QY 302 AAAATATTTTATTTTAAATAAAGCTGTGTTCAAGCTGATCATATTTCTTTATTT 361  
Db 2271 AAAATATTTTATTTTAAATAAAGCTGTGTTCAAGCTGATCATATTTCTTTATTT 2212  
QY 362 TGATTTGGGAAATAATCTGTTTCTGTATGATGAAATGCAAAATTTTGGATTTTAA 421  
Db 2211 TGATTTGGGAAATAATCTGTTTCTGTATGATGAAATGCAAAATTTTGGATTTTAA 2152  
QY 422 TCTCATTAATTTTAAATCTAATGAAATGATTAATGACATGAAAGCAACACTTAA 481  
Db 2151 TCTCATTAATTTTAAATCTAATGAAATGATTAATGACATGAAAGCAACACTTAA 2092  
QY 482 TTACTGGCAGCTGTGTCATTTGTTCTTACTAGTCTCCAAAGGAAACCTTTAA 541  
Db 2091 TTACTGGCAGCTGTGTCATTTGTTCTTACTAGTCTCCAAAGGAAACCTTTAA 2033  
QY 542 ATTGAATTTTACGAGAAATTAATCTTAAATTAATTAATTAATTAATTAATTAATTAAT 601  
Db 2032 ACTGAATCTTACGAGAAATTAATCTTAAATTAATTAATTAATTAATTAATTAATTAAT 1973  
QY 602 GTTTACATAGTCTTTGGGATTTTACTGTTCTAATTTTATTTGAAACCTCAATTTTAC 661  
Db 1972 GTTTACATAGTCTTTGGGATTTTACTGTTCTAATTTTATTTGAAACCTCAATTTTAC 1914  
QY 662 CCAGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 713  
Db 1913 CCAGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1862  
RESULT 5  
AL358233 122176 bp DNA linear HTG 10-JUL-2001  
LOCUS AL358233  
DEFINITION Homo sapiens chromosome 1 clone RP5-835A17, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 8 unordered pieces.  
ACCESSION AL358233  
VERSION AL358233.3 GI:9797852  
KEYWORDS HTG, HTGS, PHASE1, HTGS\_CANCELLED.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1  
AUTHORS McIay, K.  
TITLE Direct Submission

JOURNAL Submitted (09-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requester: clonerequest@sanger.ac.uk  
On Aug 12, 2000 this sequence version replaced gi:9214208.  
----- Genome Center

Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: dj835a17  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid: 108752; 100% of reads Chemistry:  
Dye-terminator Big Dye, 79% of reads  
Consensus quality: 11880 bases at least Q40  
Consensus quality: 120216 bases at least Q30  
Consensus quality: 120979 bases at least Q20  
Insert size: 121476; sum-of-contigs  
Insert size: 136681; 7.7% error; agarose-fp  
Quality coverage: 4.05x in Q20 bases; sum-of-contigs Quality  
coverage: 3.75x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 13486: contig of 13486 bp in length  
\* 13487 13586: gap of 100 bp  
\* 13587 30893: contig of 17307 bp in length  
\* 30894 30993: gap of 100 bp  
\* 30994 37022: contig of 6029 bp in length  
\* 37023 37122: gap of 100 bp  
\* 37123 45381: contig of 8259 bp in length  
\* 45382 45481: gap of 100 bp  
\* 45482 56788: contig of 11307 bp in length  
\* 56789 56888: gap of 100 bp  
\* 56889 84391: contig of 27503 bp in length  
\* 84392 84491: gap of 100 bp  
\* 84492 117843: contig of 33352 bp in length  
\* 117844 122176: contig of 100 bp  
\* 117944 122176: contig of 4233 bp in length.

## FEATURES

source

1.122176  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP5-835A17"  
/clone\_1fb="RPCL-5"  
1.13486  
/note="assembly\_fragment:01065  
fragment\_chain:1"  
13587..30893  
/note="assembly\_fragment:00181  
fragment\_chain:1"  
30994..37022  
/note="assembly\_fragment:00153  
fragment\_chain:1"  
37123..45381  
/note="assembly\_fragment:00580  
fragment\_chain:2"  
45482..56788  
/note="assembly\_fragment:00948  
fragment\_chain:2"  
56889..84391  
/note="assembly\_fragment:00192"  
84492..117843

/note="assembly\_fragment:00299.0"  
misc\_feature 117944..122176  
/note="assembly\_fragment:00926"  
BASE COUNT 31142 a 30481 c 29651 g 30201 t 701 others  
ORIGIN

Query Match 94.2%; Score 672.8; DB 2; Length 122176;  
Best Local Similarity 98.2%; Pred. No. 5.6e-89;  
Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 2 GCCAGAAAAAGTTATTTTATTTCTATTAACAATCTTCTCAAGCATTAATTTATCT 61  
DB 47281 GCCAGAAAAAGTTATTTTATTTCTATTAACAATCTTCTCAAGCATTAATTTATCT 47340  
QY 62 ATATCTCTGTAATTTTAAGAAATACCTTGATTTAGAAAAACCTAGAAAAAATATTA 121  
DB 47341 ATATCTCTGTAATTTTAAGAAATACCTTGATTTAGAAAAACCTAGAAAAAATATTA 47400  
QY 122 TGCAGATATTAACCTTACATGAAAAAGAAAAATTAACAAGAGCTGAGAACCTTATA 181  
DB 47401 TGCAGATATTAACCTTACATGAAAAAGAAAAATTAACAAGAGCTGAGAACCTTATA 47460  
QY 182 AATGAAATGAGATTATTAATTTGAAAACTGCATCGAAGCAAACTTATTTGTCATTA 241  
DB 47461 AATGAAATGAGATTATTAATTTGAAAACTGCATCGAAGCAAACTTATTTGTCATTA 47520  
QY 242 TNCCTTAATGATGCTGTTTATTAAGCTATACCTGATTTTTCAGAGAAACCCATGTTA 301  
DB 47521 TTCTTAATGAGGCTGTTTATTAAGCTATACCTGATTTTTCAGAGAAACCCATGTTA 47580  
QY 302 AAAATATTTTATTTTAAAAAATAGCCGTGTCAGCTGATATATTTCTTTATTT 361  
DB 47581 AAAATATTTTATTTTAAAAAATAGCCGTGTCAGCTGATATATTTCTTTATTT 47640  
QY 362 TGATTTGGAAATAATACCTGTTCTGATACATGAATGCAAAATTTTATTTTAA 421  
DB 47641 TGATTTGGAAATAATACCTGTTCTGATACATGAATGCAAAATTTTATTTTAA 47700  
QY 422 TCTCTATTTTAACTTAACTTGAATGAAATGATTAATGACATGAAGTCACACACTTA 481  
DB 47701 TCTCTATTTTAACTTAACTTGAATGAAATGATTAATGACATGAAGTCACACACTTA 47760  
QY 482 TTACGGCCAGCTGTGTCATTTGTTCTTACTAGTCCCAAGGAAACCTTAA 541  
DB 47761 TTACGGCCAGCTGTGTCATTTGTTCTTACTAGTCCCAAGGAAACCTTAA 47819  
QY 542 ATTGAATCTTACAGCAATAATCTTAATATTAATCTTTGAAGCAAAAGCTTTT 601  
DB 47820 ATTGAATCTTACAGCAATAATCTTAATATTAATCTTTGAAGCAAAAGCTTTT 47879  
QY 602 GTTACATAGTTCTTTGGATTTTACCTGTTCTTAATTTTATTTGAAATCTCAATTTACC 661  
DB 47880 GTTACATAGTTCTTTGGATTTTACCTGTTCTTAATTTTATTTGAAATCTCAATTTACC 47938  
QY 662 CCAGCCATTAATTAACATTAATCTTGTGTCAGAGCTGTTGGCAATTC 713  
DB 47939 CCAGCCATTAATTAACATTAATCTTGTGTCAGAGCTGTTGGCAATTC 47990

RESULT 6  
AC022728/c 175046 bp DNA linear HTG 24-AUG-2002  
LOCUS Homo sapiens chromosome 1 clone RP11-310J14 map 1, WORKING DRAFT  
DEFINITION  
SEQUENCE, 34 unordered pieces.  
ACCESSION  
AC022728  
VERSION  
AC022728.4 GI:7249198  
KEYWORDS  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 175046)  
Britten, B., Linton, L., Nussbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 1, clone RP11-310U14  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 175046)  
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F., Boguslavsky, L., Boucknight, B., Brown, A., Burkett, G., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArillano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Titrrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 175046)  
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boucknight, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L., Grant, P., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazars, R., Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Titrrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A., and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Mar 16, 2000 this sequence version replaced gi:6980310. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu  
Project Information

Center project name: LS491  
Center clone name: 310\_J\_14

Summary Statistics

Sequencing vector: M13: M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 156135 bases at least Q40  
Consensus quality: 165373 bases at least Q30  
Consensus quality: 168980 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 171746; sum-of-contigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 34 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 56: contig of 56 bp in length  
57 156: gap of 100 bp  
157 1191: contig of 1035 bp in length  
1192 1291: gap of 100 bp  
1292 2432: contig of 1141 bp in length  
2433 2532: gap of 100 bp  
2533 3742: contig of 1210 bp in length  
3743 3842: gap of 100 bp  
3843 4921: contig of 1079 bp in length  
4922 5021: gap of 100 bp  
5022 6073: contig of 1052 bp in length  
6074 6173: gap of 100 bp  
6174 7552: contig of 1379 bp in length  
7553 7553: gap of 100 bp  
7553 9001: contig of 1349 bp in length  
9001 9101: gap of 100 bp  
9101 10622: contig of 1521 bp in length  
10623 10722: gap of 100 bp  
10723 11836: contig of 1114 bp in length  
11837 11936: gap of 100 bp  
11937 13398: contig of 1462 bp in length  
13399 13498: gap of 100 bp  
13499 14949: contig of 1451 bp in length  
14950 15049: gap of 100 bp  
15050 16260: contig of 1211 bp in length  
16261 16360: gap of 100 bp  
16361 18265: contig of 1305 bp in length  
18266 18365: gap of 100 bp  
18366 19916: contig of 1551 bp in length  
19917 20016: gap of 100 bp  
20017 22939: contig of 2923 bp in length  
22940 23039: gap of 100 bp  
23040 26445: contig of 3406 bp in length  
26446 26545: gap of 100 bp  
26546 29327: contig of 2782 bp in length  
29328 29427: gap of 100 bp  
29428 33200: contig of 2773 bp in length  
33201 32300: gap of 100 bp  
32301 36738: contig of 4438 bp in length  
36739 36838: gap of 100 bp  
36839 43447: contig of 6609 bp in length  
43448 43547: gap of 100 bp  
43548 48525: contig of 4978 bp in length  
48526 48625: gap of 100 bp  
48626 52791: contig of 4166 bp in length  
52792 52891: gap of 100 bp  
52892 58391: contig of 5500 bp in length  
58392 58491: gap of 100 bp  
58492 64869: contig of 6378 bp in length  
64870 64969: gap of 100 bp  
64970 73369: contig of 8400 bp in length  
73370 73469: gap of 100 bp  
73470 81101: contig of 7632 bp in length  
81102 81201: gap of 100 bp  
81202 87754: contig of 6553 bp in length  
87755 87854: gap of 100 bp  
87855 95534: contig of 7680 bp in length  
95535 102797: contig of 7163 bp in length  
102798 102897: gap of 100 bp  
102899 111817: contig of 8920 bp in length  
111818 111917: gap of 100 bp  
111918 121160: contig of 9243 bp in length  
121161 121260: gap of 100 bp  
121261 140537: contig of 19277 bp in length  
140538 140637: gap of 100 bp

\* 140638..175046: contig of 34409 bp in length.

FEATURES  
Source

1..175046  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1"

/clone="RP11-310J14"  
/clone\_11b="RP11 Human Male BAC"

misc\_feature  
1..56  
/clone="RP11-310J14"  
/clone\_11b="RP11 Human Male BAC"

misc\_feature  
157..1191  
/note="assembly\_fragment"

misc\_feature  
1292..2432  
/note="assembly\_fragment"

misc\_feature  
2533..3742  
/note="assembly\_fragment"

misc\_feature  
3843..4921  
/note="assembly\_fragment"

misc\_feature  
5022..6073  
/note="assembly\_fragment"

misc\_feature  
clone\_end:SP6  
vector\_end:SP6

misc\_feature  
6174..7552  
/note="assembly\_fragment"

misc\_feature  
7653..9001  
/note="assembly\_fragment"

misc\_feature  
9102..10622  
/note="assembly\_fragment"

misc\_feature  
10723..11836  
/note="assembly\_fragment"

misc\_feature  
11937..13398  
/note="assembly\_fragment"

misc\_feature  
13489..14849  
/note="assembly\_fragment"

misc\_feature  
15050..16260  
/note="assembly\_fragment"

misc\_feature  
16361..18265  
/note="assembly\_fragment"

Query Match 94.2%; Score 672.8; DB 2; Length 175046;  
Best Local Similarity 98.2%; Pred. No. 4.9e-89;  
Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 2 GCCAGAAAAGTATTTATTTCTATTTAAACATTTCTTCAAGACATTAATTTATCCT 61  
DB 43357 GCCAGAAAAGTATTTATTTCTATTTAAACATTTCTTCAAGACATTTATTTATCCT 43298  
QY 62 ATATCTCATGAAATTTAAGAAATTAACATAGTATTAGAAAACTAGAAAAAAGATATA 121  
DB 43297 ATATCTCATGAAATTTAAGAAATTAACATAGTATTAGAAAACTAGAAAAAAGATATA 43238  
QY 122 TGCAGATATTTAACTTACATGAAAAAGAAAAATTATTAACAAGAGCTGAGAACCTTATA 181  
DB 43237 TGCAGATATTTAACTTACATGAAAAAGAAAAATTATTAACAAGAGCTGAGAACCTTATA 43178  
QY 182 AATTGAATGAGATATATTTTGAAGAACTGCACTGAAAAAGCAAACTTTATTTGTTAAATTA 241  
DB 43177 AATTGAATGAGATATATTTTGAAGAACTGCACTGAAAAAGCAAACTTTATTTGTTAAATTA 43118  
QY 242 TACTTAAATGATGCTTTTAAAGCACTAATACATGATTTTTCAGAGAGAAACCCATGTTTA 301  
DB 43117 TACTTAAATGATGCTTTTAAAGCACTAATACATGATTTTTCAGAGAGAAACCCATGTTTA 43058  
QY 302 AAAATATTTTATTTTAAATTAAGCCCTGCTGTTCAAGCTCGATCATATTTCTTTATTT 361  
DB 43057 AAAATATTTTATTTTAAATTAAGCCCTGCTGTTCAAGCTCGATCATATTTCTTTATTT 42998  
QY 362 TGATTTGGAAATAATCTGTTTCTGATAGATGAATAATGCAAAATTTTGTGATTTTAA 421

DB 42997 TGATTTGGAAATAATCTGTTTCTGATAGATGAATAATGCAAAATTTTGTGATTTTAA 42938  
QY 422 TCTACATTAATTTTAAACATTTTGAAGAAATTTGATTAATGACATGAGCCAAACACTAA 481  
DB 42937 TCTACATTAATTTTAAACATTTTGAAGAAATTTGATTAATGACATGAGCCAAACACTAA 42878  
QY 482 TTACTGGCCAGCTGTTGGCATTTGTTTCTTACTTAAGTTCGCCAAGGAAACACTTTAA 541  
DB 42877 TTACTGGCCAGCTGTTGGCATTTGTTTCTTACTTAAGTTCGCCAAGGAAACACTTTAA 42819  
QY 542 ATTGAATCTTCAGAGAAATATCTTAATATTAATTTGTAAGCAAAACAAAGCTTTT 601  
DB 42818 ACTGAATCTTCAGAGAAATATCTTAATATTAATTTGTAAGCAAAACAAAGCTTTT 42759  
QY 602 GTTACATAGTCTTTGGGATTTTACCTGTTCTTAATTTTATTCGAAACCTAATTTTACC 661  
DB 42758 GTTACATAGTCTTTGGGATTTTACCTGTTCTTAATTTTATTCGAAACCTAATTTTACC 42700  
QY 662 CCAGACCATTAATTAATTAATTTGTTTGGCAGAGTTGTTGCCAATTC 713  
DB 42698 CCAGACCATTAATTAATTAATTTGTTTGGCAGAGTTGTTGCCAATTC 42648

RESULT 7  
AL445183/c  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP11-334A14 on chromosome 1, complete  
sequence.

ACCESSION  
AL445183  
VERSION  
AL445183.19 GI:20068427  
KEYWORDS  
HTG.

SOURCE  
ORGANISM  
Homo sapiens (human)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1  
Wallis, J.  
Direct Submission  
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
On Apr 7, 2002 this sequence version replaced gi:17939714.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Wp., WormPep; Information on the WormPep  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormPep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP11-334A14 is from the library RP11-11.2 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/backpac/home.htm  
VECTOR: pBAC3.6.

FEATURES  
Source

1..193774  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

/chromosome="1"  
 /clone="RP11-334A14"  
 /clone.lib="RP11-11.2"  
 BASE COUNT 50815 a 45351 c 44031 g 53577 t  
 ORIGIN

Query Match 94.2%; Score 672.8; DB 9; Length 193774;  
 Best Local Similarity 98.2%; Pred. No. 4.7e-89;  
 Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```

QY 2 GCCAGAAAAAGTATTTTAACTTTCTATTAAACATCTCTCAAGCAATTTATTCCT 61
DB 98224 GCCAGAAAAAGTATTTTAACTTTCTATTAAACATCTCTCAAGCAATTTATTCCT 98165
QY 62 AATTCCTCACTGAATTTTAAAGAAATACATTTAGATTGAAAAAAGTAA 121
DB 98164 AATTCCTCACTGAATTTTAAAGAAATACATTTAGATTGAAAAAAGTAA 98105
QY 122 TGCAGATTAATTAACCTTACATGAAAAAGAAATTTAAACAAAGCACTGAGAACCTTATA 181
DB 98104 TGCAGATTAATTAACCTTACATGAAAAAGAAATTTAAACAAAGCACTGAGAACCTTATA 98045
QY 182 AATTGGAATGAGATTATTAATTTGAAAGCTGCACTGGAAGCAAACTTTATTTGCAATTA 241
DB 98044 AATTGGAATGAGATTATTAATTTGAAAGCTGCACTGGAAGCAAACTTTATTTGCAATTA 97985
QY 242 TTCTTAATGATGCTGTTTATGACTAATACATGATTTTTCAGAAAGAAACCCATGTTA 301
DB 97984 TTCTTAATGATGCTGTTTATGACTAATACATGATTTTTCAGAAAGAAACCCATGTTA 97925
QY 302 AAAATATTTTATTTTAAAAATAGCTGCTGTTCAAGCTCTGATCATATTTCTTTTATTT 361
DB 97924 AAAATATTTTATTTTAAAAATAGCTGCTGTTCAAGCTCTGATCATATTTCTTTTATTT 97865
QY 362 TGATTTGGGAAAAAATCTGTTTCTGATGAGCATGAATGCAAAATTTTATGATTTTAA 421
DB 97864 TGATTTGGGAAAAAATCTGTTTCTGATGAGCATGAATGCAAAATTTTATGATTTTAA 97805
QY 422 TCTCACTAATTTTAACTATTTGAGAAATGATTAATGACATGAAAGTCAACACATTA 481
DB 97804 TCTCACTAATTTTAACTATTTGAGAAATGATTAATGACATGAAAGTCAACACATTA 97745
QY 482 TTACTGGCCAGCTGTGGCATGCTGTTTCTTACTAGTTCTCCAGAGGAAACCTTTAA 541
DB 97744 TTACTGGCCAGCTGTGGCATGCTGTTTCTTACTAGTTCTCCAGAGGAAACCTTTAA 97686
QY 542 ATTGAATCTTCCAGCAATATCTTTAAATATCTTTGTAAGCAAAACAAAGCTTTT 601
DB 97685 ACTGAATCTTCCAGCAATATCTTTAAATATCTTTGTAAGCAAAACAAAGCTTTT 97626
QY 602 GTTTCATAGTCTTGGGATTTTACGTTCTTCTAATTTTATTCGAAACCTAATTTTACC 661
DB 97626 GTTTCATAGTCTTGGGATTTTACGTTCTTCTAATTTTATTCGAAACCTAATTTTACC 97567
QY 662 CCAGACCAATTAATCACTATTAATCTTTGTTGACAGCTGTTGCTCAATTC 713
DB 97566 CCAGACCAATTAATCACTATTAATCTTTGTTGACAGCTGTTGCTCAATTC 97515

```

RESULT 8  
 BC005911/c 1439 bp mRNA linear PRI 12-JUL-2001  
 LOCUS Homo sapiens, steroid carrier protein 2, clone MGC:14505  
 DEFINITION IMAGE:4287946, mRNA, complete cds.  
 ACCESSION BC005911  
 VERSION BC005911.1 GI:13543502  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1439)  
 AUTHORS Strausberg,R.

TITLE Direct Submission  
 JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcdexaxil@stanford.edu](mailto:mcdexaxil@stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>  
 Series: IMAL Plate: 21 Row: n Column: 10  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 432978.

FEATURES  
 source  
 Location/Qualifiers  
 1..1439

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="LocusID:6342"  
 /db\_xref="taxon:9606"  
 /clone="MGC:14505 IMAGE:4287946"  
 /tissue\_type="Brain, primitive neuroectodermal"  
 /clone.lib="NTH MGC\_56"  
 /lab\_host="DH10B"  
 /note="Vector: pDNR-LIB"  
 74..505  
 /codon\_start=1  
 /product="steroid carrier protein 2"  
 /protein\_id="AAH05911.1"  
 /db\_xref="GI:13543503"  
 /translation="MGFEAASFRTHQIEAVPTSSADGFKANLVEKEIEKLEEG  
 EGFVKIGIGIFAEVKGDPGKKEATWVDVNGKGSYLPSNBDKADCTITWADSDFLA  
 LMTGKNMPOSFPGKLIKITGNMELAKNLQLQPNNAKL"

BASE COUNT 490 a 219 c 272 g 458 t  
 ORIGIN

Query Match 93.5%; Score 667.6; DB 9; Length 1439;  
 Best Local Similarity 97.9%; Pred. No. 1.5e-87;  
 Matches 695; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

```

QY 4 CAGAAAAAGTATTTTAACTTTCTATTAAACATCTCTCAAGCAATTTATTCCTAT 63
DB 1409 CAGAAAAAGTATTTTAACTTTCTATTAAACATCTCTCAAGCAATTTATTCCTAT 1350
QY 64 ATCTCAGTATTTTAAAGAAATACATTTAGTATGAAAAAAGTAAATG 123
DB 1349 ATCTCAGTATTTTAAAGAAATACATTTAGTATGAAAAAAGTAAATG 1250
QY 124 CAGATTAATTAACCTTACATGAAAAAGAAATTTAAACAAGACCTAGAAAGCTTTATPAA 183
DB 1289 CAGATTAATTAACCTTACATGAAAAAGAAATTTAAACAAGACCTAGAAAGCTTTATPAA 1230
QY 184 TTGAAAGAGATTTAATTTGAAACATGCACTGGAAGCAAACTTTATTTGTTCAATTTATN 243
DB 1229 TTGAAAGAGATTTAATTTGAAACATGCACTGGAAGCAAACTTTATTTGTTCAATTTATN 1170
QY 244 CTTAATGATGCTGTTTATGACTAATACATGATTTTTCAGAAAGAAAGCAATGTTTAA 303
DB 1169 CTTAATGATGCTGTTTATGACTAATACATGATTTTTCAGAAAGAAAGCAATGTTTAA 1110
QY 304 AATATTTTATTTTAAAAATAGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTG 363
DB 1109 AATATTTTATTTTAAAAATAGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTG 1050

```

QY	375	AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTAAAGATTTTAAATCTTCACATAATTTT	434
Db	361	AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTAAAGATTTTAAATCTTCACATAATTTT	420
QY	435	AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTAAAGATTTTAAATCTTCACATAATTTT	494
Db	421	AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTAAAGATTTTAAATCTTCACATAATTTT	480
QY	495	GTGGCATGTGTTCTGATAGCATGAAATGCAAAATTTTAAAGATTTTAAATCTTCACATAATTTT	554
Db	481	GTGGCATGTGTTCTGATAGCATGAAATGCAAAATTTTAAAGATTTTAAATCTTCACATAATTTT	539
QY	555	CAGATTAATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	614
Db	540	CAGATTAATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	597
QY	615	TTTGGGATTTTACGTTCTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	674
Db	598	TTTGGGATTTTACGTTCTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	656
QY	675	ACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	704
Db	657	ACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	687
RESULT 10			
LOCUS	BD079829	687 bp DNA linear	PAT 27-AUG-2002
DEFINITION	Cancer-associated nucleic acids and polypeptides.		
VERSION	BD079829.1	GI:22625432	
KEYWORDS	JP 2001516009-A/495.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 687)		
TITLE	Old, L. J., Scanlan, M. J., Stockert, E., Gure, A., Chen, Y. T., Gout, I., O'Gare, M., Obara, Y., Preundschuh, M., Tureci, O. and Sahin, U.		
JOURNAL	Cancer-associated nucleic acids and polypeptides		
COMMENT	Patent: JP 2001516009-A 495 25-SEP-2001; LUDWIG INSTITUTE FOR CANCER RESEARCH		
	OS Homo sapiens (human)		
	PN JP 2001516009-A/495		
	PD 25-SEP-2001		
	PR 15-JUL-1998 JP 2000503425		
	PR 17-JUL-1997 US 08/896164, 10-OCT-1997 US 60/061599 PR		
	10-OCT-1997 US 60/061765, 10-OCT-1997 US 08/948705 PR		
	11-OCT-1997 GB 9721697.2, 22-JUN-1998 US 09/102322 PI		
	J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI TSENG		
	CHEN, PI IVAN GOUT, MICHAEL O'HARE, YUICHI OBATA, MICHAEL PREUNDSCUH, PI		
	OZLEM TURECI,		
	PI UGUR SAHIN		
	PC		
	GOIN33/574, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61K48/00, PC		
	A61P35/00,		
	PC C07K14/82, C07K16/32, C12N15/09//C07K16/46, C12P21/08, A61K37/02,		
	PC C12N15/00		
	CC Cancer-associated nucleic acids and polypeptides. FH Key		
	Location/Qualifiers		
	FT source 1..687		
	FT Location/Qualifiers		
	1..687 /organism="Homo sapiens (human)".		
FEATURES	Location/Qualifiers		
SOURCE	1..687		
BASE COUNT	242 a 242 c 85 g 258 t 4 others		
ORIGIN			
Query Match	85.7%, Score 611.8, DB 6, Length 687,		

Best Local Similarity 97.1%; Pred. No. 2.3e-79;  
Matches 671; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

```
OY 15 ATTTAATTTTCTATTAACATTTCTTCAAGCATATTTATCTCACTGAA 74
Db 1 ATTTAATTTTCTATTAACATTTCTTCAAGCATATTTATCTCACTGAA 60
OY 75 TTTTAAAGAAATACATTTAGTATTTAGAAAACTAGGAAAAAAGATTAATSCAGATATTTA 134
Db 61 TTTTAAAGAAATACATTTAGTATTTAGAAAACTAGGAAAAAAGATTAATSCAGATATTTA 120
OY 135 ACTTACATGAAAAAGAAAAATTATACAAAGAGCTGAGACCTTATTAATTTGAAATGAGA 194
Db 121 ACTTACATGAAAAAGAAAAATTATACAAAGAGCTGAGACCTTATTAATTTGAAATGAGA 180
OY 195 TTTAATATTTGAAAACTGCACTGGAAGAGCAACTTATTTCTCAATTATTTCTAATGATGG 254
Db 181 TTTAATATTTGAAAACTGCACTGGAAGAGCAACTTATTTCTCAATTATTTCTAATGATGG 240
OY 255 TGTTTATGACTAATACACTGATTTTTTCAAGAGAAACCATGTTAAATAATTTTAT 314
Db 241 TGTTTATGACTAATACACTGATTTTTTCAAGAGAAACCATGTTAAATAATTTTAT 300
OY 315 TTTAAAAATTAAGCCCTGTGTTCAGCTCTGATCATATTTCTTTATTTTGTATTTGGAGANA 374
Db 301 TTTAAAAATTAAGCCCTGTGTTCAGCTCTGATCATATTTCTTTATTTTGTATTTGGAGANA 360
OY 375 AAATCTGTTTCTGATAGATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 434
Db 361 AAATCTGTTTCTGATAGATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 420
OY 435 AANAATCTATGAGATTTGATTAATGACATGAAGAGCAACACTAATTTACTGGCAGCT 494
Db 421 AAGAATCTATGAGATTTGATTAATGACATGAAGAGCAACACTAATTTACTGGCAGCT 480
OY 495 GTTGCAATTTGTTTCTTACTAGTTCTCCCAAGGAAAACTTTAAATTTGATCTTCAAG 554
Db 481 GTTGCAATTTGTTTCTTACTAGTTCTCCCAAGGAAAACTTTAAATTTGATCTTCAAG 539
OY 555 CAGATTAATCTTAAATATATCTTTGTAGCAAAACAAAGCTTTTGTTCATATGATTC 614
Db 540 CAGATTAATCTTAAATATATCTTTGTAGCAAAACAAAGCTTTTGTTCATATGATTC 597
OY 615 TTTGGATTTTACCTGTTCTTATTTTATTTCTGAACCTCAATTTTACCAGCAGCAATTT 674
Db 588 TTT-GGATTTTACCTGTTCTTATTTTATTTCTGAACCTCAATTTTACCAGCAGCAATTT 656
OY 675 ACCATA-TTAACTTTGTTNCGACAGTTGTT 704
Db 657 ACCCATATTAATTTGTTATGACAGTTGTT 687
```

RESULT 11  
S52450 1229 bp mRNA linear PRI 08-MAY-1993  
LOCUS sterol carrier protein 2 [human, liver, mRNA, 1229 nt].  
DEFINITION S52450  
ACCESSION S52450  
VERSION S52450.1 GI:263550  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1229)  
AUTHORS Yamamoto, R.  
TITLE Localization of human sterol carrier protein 2 gene and cDNA  
expression in COS-7 cell  
JOURNAL Hokkaido Igaku Zasshi 67 (6), 839-848 (1992)  
MEDLINE 93331254  
PUBMED 1483685  
REMARK GenBank staff at the National Library of Medicine created this  
entry [NCBI g122904] from the original journal article.  
This sequence comes from Fig. 1.

Map location: 1.  
FEATURES Location/Qualifiers

source 1..1229  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

gene 1..1229  
/gene="sterol carrier protein 2, SCP2"  
84..515  
/gene="sterol carrier protein 2, SCP2"  
/note="This sequence comes from Fig. 1; SCP2"  
/codon\_start=1  
/product="sterol carrier protein 2"  
/protein\_id="AAB24921.1"  
/db\_xref="GI:263551"

CDS  
BASE COUNT 409 a 200 c 255 g 365 t

ORIGIN

Query Match 66.5%; Score 474.6; DB 9; Length 1229;  
Best Local Similarity 97.1%; Pred. No. 1.2e-59;  
Matches 502; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

```
OY 195 TTTAATTTGAAACCTGATCTGAAGCAACTTTATTTGTTCAATTTATTTATGATGG 254
Db 1228 TTTTATTTGAAACCTGATCTGAAGCAACTTTATTTGTTCAATTTATTTATGATGG 1169
OY 255 TGTTTATGACTAATACACTGATTTTTTCAAGAGAAACCATGTTAAATAATTTTAT 314
Db 1168 TGTTTATGACTAATACACTGATTTTTTCAAGAGAAACCATGTTAAATAATTTTAT 1109
OY 315 TTTAAAAATTAAGCCCTGTGTTCAGCTCTGATCATATTTCTTTATTTTGTATTTGGAGANA 374
Db 1108 TTTAAAAATTAAGCCCTGTGTTCAGCTCTGATCATATTTCTTTATTTTGTATTTGGAGANA 1049
OY 375 AAATCTGTTTCTGATAGATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 434
Db 1048 AAATCTGTTTCTGATAGATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 989
OY 435 AANAATCTATGAGATTTGATTAATGACATGAAGAGCAACACTAATTTACTGGCAGCT 494
Db 988 AAGAATCTATGAGATTTGATTAATGACATGAAGAGCAACACTAATTTACTGGCAGCT 929
OY 495 GTTGCAATTTGTTTCTTACTAGTTCTCCCAAGGAAAACTTTAAATTTGATCTTCAAG 554
Db 928 GTTGCAATTTGTTTCTTACTAGTTCTCCCAAGGAAAACTTTAAATTTGATCTTCAAG 870
OY 555 CAGATTAATCTTAAATATATCTTTGTAGCAAAACAAAGCTTTTGTTCATATGATTC 614
Db 869 CAGATTAATCTTAAATATATCTTTGTAGCAAAACAAAGCTTTTGTTCATATGATTC 810
OY 615 TTTGGATTTTACCTGTTCTTATTTTATTTCTGAACCTCAATTTTACCAGCAGCAATTT 674
Db 809 TTT-GGATTTTACCTGTTCTTATTTTATTTCTGAACCTCAATTTTACCAGCAGCAATTT 751
OY 675 ACCATATTAATTTGTTNCGACAGTTGTTGCAAT 711
Db 750 ACCATATTAATTTGTTNCGACAGTTGTTGCAAT 714
```

RESULT 12  
LOCUS HUMSTEAALC 1219 bp mRNA linear PRI 06-DEC-1993  
DEFINITION Human sterol carrier protein-2 (SCP-2) mRNA, complete cds.  
ACCESSION M55421  
VERSION M55421.1 GI:432978  
KEYWORDS sterol carrier protein-2.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1219)  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE Localization of human sterol carrier protein 2 gene and cDNA  
expression in COS-7 cell  
JOURNAL Hokkaido Igaku Zasshi 67 (6), 839-848 (1992)  
MEDLINE 93331254  
PUBMED 1483685  
REMARK GenBank staff at the National Library of Medicine created this  
entry [NCBI g122904] from the original journal article.  
This sequence comes from Fig. 1.



REFERENCE 1 (bases 1 to 1219)  
AUTHORS Yamamoto,R., Kallen,C.B., Babalola,G.O., Rennett,H.,  
TITLE Billheimer,J.T. and Straus,J.F. III  
Cloning and expression of a cDNA encoding human sterol carrier  
protein 2  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (2), 463-467 (1991)  
MEDLINE 9110550  
PUBMED 1703300  
REFERENCE 2 (bases 1 to 1219)  
AUTHORS Vesa,J., Hellsten,E., Straus,J.F. III, and Peltonen,L.  
TITLE Mead,S., Cowell,J.K., Straus,J.F. III, and Peltonen,L.  
Assignment of sterol carrier protein X/sterol carrier protein 2 to  
1p32 and exclusion as the causative gene for infantile neuronal  
ceroid lipofuscinosis  
JOURNAL Unpublished  
COMMENT On Dec 6, 1993 this sequence version replaced gi:432972.  
FEATURES  
source location/Qualifiers  
1..1219  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="1p32"  
/tissue\_type="liver"  
1..1215  
/gene="SCP-2"  
84..515  
/gene="SCP-2"  
/codon\_start=1  
/product="sterol carrier protein-2"  
/protein\_id="AAA0359.1"  
/db\_xref="GI:432973"  
/translation="MGFPEAASSFRTTHQIEAVPTSSADGKRAIVFEIKKEEES  
EQFVKIKGIFAFKVDGPGKEATVVDVKNKGSLVPSDKADCTTWDSDFLA  
LMTGKNPQSAFFQCKLKITGNMGLAMKLQLOPGNAKL"  
84..143  
/gene="SCP-2"  
144..512  
/gene="SCP-2"  
/product="unnamed"  
782..787  
/gene="SCP-2"  
/evidence=not\_experimental  
polyA\_signal  
1062..1067  
/gene="SCP-2"  
/evidence=not\_experimental  
polyA\_signal  
1107..1112  
/gene="SCP-2"  
/evidence=not\_experimental  
polyA\_signal  
1190..1195  
/gene="SCP-2"  
/evidence=not\_experimental  
polyA\_signal  
1219  
/gene="SCP-2"  
/evidence=experimental  
polyA\_site  
/evidence=experimental  
BASE COUNT 402 a 195 c 255 g 367 t  
ORIGIN  
Query Match 64.4%; Score 459.8; DB 9; Length 1219;  
Best Local Similarity 97.3%; Pred. No. 1.7e-57;  
Matches 497; Conservative 0; Mismatches 11; Indels 3; Gaps 3;

QY 383 TTTCGATAGCATGAAGAAATGCAAAATTTTAACTTTTAATCTCACTAATTTTAAATGA 442  
DB 1039 TTTCGATAGCATGAAGAAATGCAAAATTTTAACTTTTAACTAATTTTAAAGACTA 981  
QY 443 TTGAGAAATTTGATTAATGACATGAAAGTGACACAACCTAATTTACTGGCCAGCTGTGGCAT 502  
DB 980 TTGAGAAATTTGATTAATGACATGAAAGTGACACAACCTAATTTACTGGCCAGCTGTGGCAT 921  
QY 503 TGTGTTTCTTACTTAACTTCTCCCAAGGAAATCTTAAATTTGAATCTTCACGAGAAATGA 562  
DB 920 TGTGTTTCTTACTTAACTTCTCCCAAGGAAATCTTAAATTTGAATCTTCACGAGAAATGA 862  
QY 563 TCCTTAAATATACCTTTTGACCAAAACAAAGCTTTTGTTCATAGTCTTTGGAT 622  
DB 861 TCCTTAAATATACCTTTTGACCAAAACAAAGCTTTTGTTCATAGTCTTTGGAT 803  
QY 623 TTTCATGTTCCCTAATTTTATTTCTGAATCTCAATTTTACCAGACCATTAATTTACCATATT 682  
DB 802 TTTCATGTTCCCTAATTTTATTTCTGAATCTCAATTTTACCAGACCATTAATTTACCATATT 743  
QY 683 AACTTGTGTTGACACAGTTGTTGCCAATTC 713  
DB 742 AACTTGTGTTGACACAGTTGTTGCCAATTC 712  
RESULT 13  
ARI46582  
LOCUS ARI46582 843 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 22 from patent US 6218521.  
ACCESSION ARI46582  
VERSION ARI46582.1 GI:15109771  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 843)  
AUTHORS Obata,Y.  
TITLE Isolated nucleic acid molecules associated with gastric cancer and  
methods for diagnosing and treating gastric cancer  
JOURNAL Patent: US 6218521-A 22 17-Apr-2001;  
FEATURES  
source location/Qualifiers  
1..843  
/organism="unknown"  
BASE COUNT 237 a 184 c 89 g 215 t 118 others  
ORIGIN  
Query Match 61.1%; Score 436.2; DB 6; Length 843;  
Best Local Similarity 73.1%; Pred. No. 5e-54;  
Matches 516; Conservative 0; Mismatches 187; Indels 3; Gaps 3;

Db 302 AAANNTTTTNTATTTAAAAATAACCCNGTNTCCAAACCCNGATCANATTCCTTNAATTT 361  
 QY 362 TGAATTTGGGAAANAAATAGTCTTGATAGCATGAANAATCAATTTTATGATTTTAA 421  
 Db 362 GGAATGGGGGAAAAAATNCNGTTCNNATACGNGAANNGCAANNTTTTAAATTTTAA 421  
 QY 422 TCTCATTAATTTTAAANAATTAATGAGAAATGATTAATGACATGAAGTGACAAACTAA 481  
 Db 422 CCCCCCTAATTTTAAANATGAAANNTGATTTNNGACTGAATGTC-CAACCTTAN 480  
 QY 482 TTACTGGCAGCTGGGATGTGTCTTCTACTGTCTCCAGGAAAACTCTTAA 541  
 Db 481 TTNCNGGCCACCGTGGCGCTNGTNTTCTTACTTANCCCCCAA-GGAANNCCTTAA 539  
 QY 542 ATGAATCTTCAGCAGAAATATCTTAAATATATCTTGTAGCAAAACAAAGCTTTT 601  
 Db 540 NCGAANCTCCNCAAAATATACCTTAATATCTTGTAGCAAAACAAAGCTTTT 599  
 QY 602 GTTTACATAGTCTTTGGATTTTACTGTCTTCTAATTTATCTGAACTCAATTTTAC 661  
 Db 600 GTTTACNTA-NTCCTTGGATTTAAGGGTCCCAATTTNATCCNCAACCAATTTTCC 658  
 QY 662 CCAGACCAATTAATCAATATTAATCTTGTNTGCAAGTTGTTGC 707  
 Db 659 CCNACCATANTTACATTTTACTTGTGTAGGCGCAGTNGTTGC 704

RESULT 14  
 BD079831 843 bp DNA linear PAT 27-AUG-2002  
 LOCUS BD079831  
 DEFINITION Cancer-associated nucleic acids and polypeptides.  
 ACCESSION BD079831  
 VERSION BD079831.1 GI:22625434  
 KEYWORDS JP 2001516009-A/497.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 TITLE Old, L.J., Scanlan, M.J., Stockert, E., Gure, A., Chen, Y.T., Gout, I., Ognare, M., Obata, Y., Pfeundschuh, M., Tureci, O. and Sahin, U.  
 JOURNAL Cancer-associated nucleic acids and polypeptides  
 COMMENT Patient: JP 2001516009-A 497 25-SEP-2001;  
 LUDWIG INSTITUTE FOR CANCER RESEARCH  
 OS Homo sapiens (human)  
 PN JP 2001516009-A/497  
 PD 25-SEP-2001  
 PR 15-JUL-1998 JP 2000503425  
 PR 17-JUL-1997 US 08/896164, 10-OCT-1997 US 60/061599 PR  
 11-OCT-1997 US 60/061765, 10-OCT-1997 US 08/948705 PR  
 11-OCT-1997 GB 9721697.2, 22-JUN-1998 US 09/102322 PI LLOYD  
 J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI TSENG  
 CHEN,  
 PI IVAN GOUT, MICHAEL O'HARE, YUICHI OBATA, MICHAEL PFEUNDSCHEH, PI  
 OZLEM TURECI,  
 PI UGUR SAHIN

PC G01N33/574, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61K48/00, PC  
 A61P35/00,  
 PC C07K14/82, C07K16/32, C12N15/09//C07K16/46, C12P21/08, A61K37/02,  
 PC C12N15/00  
 CC Cancer-associated nucleic acids and polypeptides. FH Key  
 Location/Qualifiers  
 FT source 1..843  
 Location/Qualifiers  
 1..843  
 Location/Qualifiers  
 /organism="Homo sapiens (human)"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 237 a 184 c 89 g 215 t 118 others  
 ORIGIN

Query Match 61.1%; Score 436.2; DB 6; Length 843;  
 Best Local Similarity 73.1%; Pred. No. 56-54;  
 Matches 516; Conservative 0; Mismatches 187; Indels 3; Gaps 3;

QY 2 GCCAGAAAAAGTTATTTTATTTCTATTTAAACATTTCTTCAAGCATTTATTCCT 61  
 Db 2 GCCAAAAAANATTTATTTNATTCCTATTTAANCTCTCCNCAANCAATTTATNACCT 61  
 QY 62 ATATCTCATGTAATTTTAAAGAAATTAACATTTAGTATTAAGAAAAATAGAAATAA 121  
 Db 62 ATTNCKCNCNGANTTTNANAANATCTTNTNTTTAAAAAAGCTTNGAAAAAATATAT 121  
 QY 122 TCCAGATTAATTAACCTTATCATGAAAAAGAAATTTATTAACAAGAGCTGAGAACCTTAT 181  
 Db 122 NCGAATANTTTAATCTTCTTGAAGAAAGAAATTTTACAGAGAAAGAAAGAAAGCTTAT 181  
 QY 182 AATGAATATGAGATTAATTTGAAAGAAAGCTGATGAAAGCAAACTTTATTTGTAATTA 241  
 Db 182 ATTNCAANTNAAATATATNTTNGAAGGCGCNGAAGCAACCAANTTTATGTCATTA 241  
 QY 242 TNCCTTAATGAGCTGTTTATGACATTAATACATGATTTTCAAGAGAAAGCAATGTTA 301  
 Db 242 TCTTNAAGAGGANTTTNANNATTAATNCCNGATTTTCAATNNGAANCCNNNTTA 301  
 QY 302 AAATATATTTTATTTTAAAAATTAAGCTGTGTCAAGCTGATCATATTTCTTTATTT 361  
 Db 302 AAANNTTTTATTTTAAAAATTAACCCNGTNTCCAAACCCNGATCANATTCCTTNAATTT 361  
 QY 362 TGAATTTGGGAAANAAATCTGTTTCTGTATGATGAAATGCAAAATTTTGAATTTTAA 421  
 Db 362 GGAATGGGGGAAAAAATNCNGTTCNNATATACCCNNGAANNGCAANNTTTTAAATTTTAA 421  
 QY 422 TCTCATTAATTTTAAANAATTAATGAGAAATGATTAATGACATGAAGTGCAACACTTA 481  
 Db 422 CCCCCCTAATTTTAAANATGAAANNTGATTTNNGACTGAATGTC-CAACCTTAN 480  
 QY 482 TTACTGGCAGCTGTGTCATTTGTCTTACTTACTTGTCTCCAAAGGAAAACTCTTAA 541  
 Db 481 TTNCNGGCCACCGTGGCGCTNGTNTTCTTACTTANTCCCCCAA-GGAANNCCTTAA 539  
 QY 542 ATGAATCTTCAGCAGAAATATCTTAAATATATCTTGTAGCAAAACAAAGCTTTT 601  
 Db 540 NCGAANCTCCNCAAAATATACCTTAATATCTTGTAGCAAAACAAAGCTTTT 599  
 QY 602 GTTTACATAGTCTTTGGATTTTACTGTCTTCTAATTTATCTGAACTCAATTTTAC 661  
 Db 600 GTTTACNTA-NTCCTTGGATTTAAGGGTCCCAATTTNATCCNCAACCAATTTTCC 658  
 QY 662 CCAGACCAATTAATCAATATTAATCTTGTNTGCAAGTTGTTGC 707  
 Db 659 CCNACCATANTTACATTTTACTTGTGTAGGCGCAGTNGTTGC 704

RESULT 15  
 HUMSCP2B/1500 bp mRNA linear PRI 06-DEC-1993  
 LOCUS HUMSCP2B  
 DEFINITION Human sterol carrier protein 2 mRNA, complete cds.  
 ACCESSION M75884  
 VERSION M75884.1 GI:432976  
 KEYWORDS sterol carrier protein-2.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 TITLE He, Z., Yamamoto, R., Futh, E.E., Schantz, L.J., Navjot, S.L., George, H., Billheimer, J.T. and Struss, J.F., III.  
 REFERENCES CDNA encoding members of a family of proteins related to human sterol carrier protein 2 and assignment of the gene to human chromosome 1 p21---pter  
 JOURNAL DNA Cell Biol. 10 (8), 559-569 (1991)  
 MEDLINE 92029618  
 PUBMED 1718316

REFERENCE  
2 (bases 1 to 1500)  
AUTHORS  
Vera, J., Hellsten, E., Branstetter, B. S., Billheimer, J. T.,  
Mead, S., Cowell, J. K., Strauss, J. F. III, and Peltonen, L.  
TITLE  
Assignment of steroid carrier protein 2 to  
1p32 and exclusion as the causative gene for infantile neuronal  
ceroid lipofuscinosis  
JOURNAL  
unpublished  
COMMENT  
On Dec 6, 1993 this sequence version replaced gi:337996.  
Original source text: Human liver cDNA to mRNA.

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:55:53 ; Search time 1924.36 Seconds  
(without alignments)  
9017.749 Million cell updates/sec

Title: US-09-835-992a-19

Perfect score: 714  
Sequence: 1 cgcacgaagaagatcttca.....cagcttcttgcacatca 714

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_estbda: \*  
2: em\_estbhm: \*  
3: em\_estcin: \*  
4: em\_estcma: \*  
5: em\_estcov: \*  
6: em\_estcpl: \*  
7: em\_estcro: \*  
8: em\_estcrl: \*  
9: gb\_estc1: \*  
10: gb\_estc2: \*  
11: gb\_estc3: \*  
12: gb\_estc4: \*  
13: gb\_estc5: \*  
14: gb\_estc6: \*  
15: em\_estfun: \*  
16: em\_estgom: \*  
17: em\_ges\_hum: \*  
18: em\_ges\_inv: \*  
19: em\_ges\_pln: \*  
20: em\_ges\_vrc: \*  
21: em\_ges\_fun: \*  
22: em\_ges\_mam: \*  
23: em\_ges\_mus: \*  
24: em\_ges\_pro: \*  
25: em\_ges\_rnd: \*  
26: em\_ges\_vhg: \*  
27: em\_ges\_vrl: \*  
28: gb\_ges1: \*  
29: gb\_ges2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	673	94.3	769	12	BQ014192
2	672.8	94.2	12	BM997078	BM997078 UI-H-ED1- A1640146 wa29f11.x
3	671.8	94.1	821	9	A1640146 wa29f11.x
4	670.4	93.9	770	9	A1826287 wk33f07.x

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H-ED1-axs-g-24-0-UI-81 NCI_CGAP_Ed1 Homo sapiens CDNA clone	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		IMAGE:5833007 3', mRNA sequence.	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		GI:19739093	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		769 bp mRNA linear EST 26-MAR-2002	BQ014192				Homo sapiens (human)	1 (bases 1 to 769)			
BQ014192		UI-H									

Seq primer: M13 FORWARD  
POLYA=Yes.

Location/Qualifiers

1..769

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5831007"  
/tissue\_type="Chondrosarcoma"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP ED1"  
/note="Organ: Left Pubic Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP ED1 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma cell line C85. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAGGCT.  
TAG\_Lib=UI-H-ED1  
TAG\_Tissue=chondrosarcoma  
TAG\_SEQ=CTCAAGCT"

BASE COUNT 272 a 110 g 285 t 2 others

ORIGIN

Query Match 94.3%; Score 673; DB 12; Length 769;  
Best Local Similarity 98.3%; Pred. No. 7e-70;  
Matches 698; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

QY 2 GCCAGAAAAAGTTATTTTATTTCTATTAACATTTCTTCAAGCATTTATTTATCTT 61  
DB 20 GCCAGAAAAAGTTATTTTATTTCTATTAACATTTCTTCAAGCATTTATTTATCTT 79

QY 62 ATATCTCATGATTTTATTAAGAAATTAACATTTATTAAGAAAACTATGAGAAAAAGATATA 121  
DB 80 ATATCTCATGATTTTATTAAGAAATTAACATTTATTAAGAAAACTATGAGAAAAAGATATA 139

QY 122 TGCAGATTAATTAACATTAAGAAAGAAATTTATTAACAAAGACTGAGAACGTTATA 181  
DB 140 TGCAGATTAATTAACATTAAGAAAGAAATTTATTAACAAAGACTGAGAACGTTATA 199

QY 182 AATTGAATGAGATTAATTTGAAAACTGATCTGAAGCAACTTATTTGTTCAATTA 241  
DB 200 AATTGAATGAGATTAATTTGAAAACTGATCTGAAGCAACTTATTTGTTCAATTA 259

QY 242 TMTTAAATGATGTTTATGACTAATACATGATTTTTCAGAGAGAAACCCATGTTA 301  
DB 260 TMTTAAATGATGTTTATGACTAATACATGATTTTTCAGAGAGAAACCCATGTTA 319

QY 302 AAAATATTTTATTTTAAAAAAGCTGCTGTTAAAGCTGATCATATTTCTTTATTT 361  
DB 320 AAAATATTTTATTTTAAAAAAGCTGCTGTTAAAGCTGATCATATTTCTTTATTT 379

QY 362 TGAATTTGGAANAATACTGTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 421  
DB 380 TGAATTTGGAANAATACTGTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 439

QY 422 TCTCTAATTTTAAANAATACTGATTAATGAAATGATTAATGATGACAACTATA 481  
DB 440 TCTCTAATTTTAAANAATACTGATTAATGAAATGATTAATGATGACAACTATA 499

QY 482 TTACTGGCCAGCTGTGGCATGTGTTTCTTACTTAAGTTCTCCAGGAAAACTCTTA 541  
DB 500 TTACTGGCCAGCTGTGGCATGTGTTTCTTACTTAAGTTCTCCAGGAAAACTCTTA 558

QY 542 ATTGAATCTTCAGAGAAATATCTTAATATATCTTTGTAGAGCAAAACAAAGCTTTT 601

Db 559 ACTGAATCTTCAGAGAAATATCTTAATATATCTTTGTAGCAAAACAAAGCTTTT 618

QY 602 GTTACATAGTTCTTTGGGATTTTACTGTTCCATATTTTATCTTGAAGCATATTTACC 661

Db 619 GTTACATAGTTCTTT-GGATTTACTGTTCCATATTTTATCTTGAAGCATATTTACC 677

QY 662 CCACACATATATTCATATTAATCTTGTGTTGACAGTTGTTGCCAT 711

Db 678 CCAGACATATATTAACATATTAATCTTGTATGACAGTTGATGCAAT 727

RESULT 2  
BM997078

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Robert Strausberg, Ph. D.  
Email: cgabps-remail.nih.gov  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: <http://image.llnl.gov>  
The following repetitive elements were found in this cDNA sequence: 316-342, >AT-richlow\_complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

Location/Qualifiers

1..761

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5831426"  
/tissue\_type="Chondrosarcoma"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP ED0"  
/note="Organ: Left Pubic Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP ED0 is a cDNA library containing the following tissue(s): Chondrosarcoma cell line C85. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAGGCT.  
TAG\_Lib=UI-H-ED0  
TAG\_Tissue=chondrosarcoma  
TAG\_SEQ=CTCAAGCT"

BASE COUNT 269 a 108 c 98 g 285 t 1 others

ORIGIN

Query Match 94.2%; Score 672.8; DB 12; Length 761;  
Best Local Similarity 98.2%; Pred. No. 7.4e-70;

Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 2 GCCGAGAAAAAGTTTAAATTTTCTATTAAACCTTCTTCGCAAGCATTTATTTACCT 61  
 DB 19 GCCGAGAAAAAGTTTAAATTTTCTATTAAACCTTCTTCGCAAGCATTTATTTATCCT 78

QY 62 ATATCTGAGTATTTTAAAGAAATACATTAATTAAGAAAACTAGAAAAAAGATATA 121  
 DB 79 ATATCTGAGTATTTTAAAGAAATACATTAATTAAGAAAACTAGAAAAAAGATATA 138

QY 122 TGCAGATTAATTAACCTACATGAAGAAAAATTAACCAAGAGCTAGAACGTTATA 181  
 DB 139 TGCAGATTAATTAACCTACATGAAGAAAAATTAACCAAGAGCTAGAACGTTATA 198

QY 182 AATTGAAATGAGATTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGCAATTA 241  
 DB 199 AATTGAAATGAGATTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGCAATTA 258

QY 242 TNCCTAATGATGGTGTATTGACTAATACATGATTTTTCAGAGAGAAACCATGTTA 301  
 DB 259 TTTCTTAATGATGGTGTATTGACTAATACATGATTTTTCAGAGAGAAACCATGTTA 318

QY 302 AAAATATTTTAAATTTTAAAGAAATAGCTGTGTCAGCTGATCATATTTCTTTATTT 361  
 DB 319 AAAATATTTTAAATTTTAAAGAAATAGCTGTGTCAGCTGATCATATTTCTTTATTT 378

QY 362 TGATTTGGGAAANAAATACCTGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 421  
 DB 379 TGATTTGGGAAANAAATACCTGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 438

QY 422 TTTCCACTAATTTTAAANAACTATGAGAAATGATTAATGACATGAAGGCAACACTTAA 481  
 DB 439 TTTCCACTAATTTTAAANAACTATGAGAAATGATTAATGACATGAAGGCAACACTTAA 498

QY 482 TTACTGGCCAGCTGTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACCTCTTAA 541  
 DB 499 TTACTGGCCAGCTGTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACCTCTTAA 557

QY 542 ATTTGAATCTTCAGAGAAATATCTTTAATATACCTTTGTAAAGCAAAACAAAGCTTTT 601  
 DB 558 ATTTGAATCTTCAGAGAAATATCTTTAATATACCTTTGTAAAGCAAAACAAAGCTTTT 617

QY 602 GTTTACATAGTCTTTGGGATTTTACTGTTCCCTAATTTTATTCGAAACCAATTTTACC 661  
 DB 618 GTTTACATAGTCTTTGGGATTTTACTGTTCCCTAATTTTATTCGAAACCAATTTTACC 676

QY 662 CCAGACCATATTAATCAATTAATTAATTTGATGACAGATTGTCGAATTC 713  
 DB 677 CCAGACCATATTAATCAATTAATTAATTTGATGACAGATTGTCGAATTC 728

RESULT 3  
 A1640146 821 bp mRNA linear EST 16-DEC-1999  
 LOCUS w629f11.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:229553 3'  
 DEFINITION similar to gp:S52450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR  
 (HUMAN); mRNA sequence.  
 A1640146  
 VERSION A1640146.1 GI:4703255  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 821)  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Streusberg, Ph.D.  
 Email: [cgapdb-roman1.nih.gov](mailto:cgapdb-roman1.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone Distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.1nl.gov/biopr/image/image.html](http://www-bio.1nl.gov/biopr/image/image.html)  
 Insert Length: 565 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 471.  
 Location/Qualifiers  
 1. 821  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:229553"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Kid11"  
 /note="Organ: Kidney; Vector: pRTT3D-Pac (Pharmacia) with  
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
 Plasmid DNA from the normalized library NCI CGAP Kid1 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneids 1322376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Fatima Bonaldo."

BASE COUNT 296 a 128 c 114 g 278 t 5 others  
 ORIGIN

Query Match 94.1%; Score 671.8; DB 9; Length 821;  
 Best Local Similarity 96.9%; Pred. No. 9.4e-70;  
 Matches 699; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

QY 2 GCCGAGAAAAAGTTTAAATTTTCTATTAAACCTTCTTCGCAAGCATTTATTTACCT 61  
 DB 19 GCCGAGAAAAAGTTTAAATTTTCTATTAAACCTTCTTCGCAAGCATTTATTTATCCT 78

QY 62 ATATCTGAGTATTTTAAAGAAATACATTAATTAAGAAAACTAGAAAAAAGATATA 121  
 DB 79 ATATCTGAGTATTTTAAAGAAATACATTAATTAAGAAAACTAGAAAAAAGATATA 138

QY 122 TGCAGATTAATTAACCTACATGAAGAAAAATTAACCAAGAGCTAGAACGTTATA 181  
 DB 139 TGCAGATTAATTAACCTACATGAAGAAAAATTAACCAAGAGCTAGAACGTTATA 198

QY 182 AATTGAAATGAGATTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGCAATTA 241  
 DB 199 AATTGAAATGAGATTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGCAATTA 258

QY 242 TNCCTAATGATGGTGTATTGACTAATACATGATTTTTCAGAGAGAAACCATGTTA 301  
 DB 259 TTTCTTAATGATGGTGTATTGACTAATACATGATTTTTCAGAGAGAAACCATGTTA 318

QY 302 AAAATATTTTAAATTTTAAAGAAATAGCTGTGTCAGCTGATCATATTTCTTTATTT 361  
 DB 319 AAAATATTTTAAATTTTAAAGAAATAGCTGTGTCAGCTGATCATATTTCTTTATTT 378

QY 362 TGATTTGGGAAANAAATACCTGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 421  
 DB 379 TGATTTGGGAAANAAATACCTGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 438

QY 422 TTTCCACTAATTTTAAANAACTATGAGAAATGATTAATGACATGAAGGCAACACTTAA 481  
 DB 439 TTTCCACTAATTTTAAANAACTATGAGAAATGATTAATGACATGAAGGCAACACTTAA 498

QY 482 TTACTGGCCAGCTGTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACCTCTTAA 541  
 DB 499 TTACTGGCCAGCTGTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACCTCTTAA 557

QY 542 ATTTGAATCTTCAGAGAAATATCTTTAATATACCTTTGTAAAGCAAAACAAAGCTTTT 601  
 DB 558 ATTTGAATCTTCAGAGAAATATCTTTAATATACCTTTGTAAAGCAAAACAAAGCTTTT 617

QY 602 GTTTACATAGTCTTTGGGATTTTACTGTTCCCTAATTTTATTCGAAACCAATTTTACC 661  
 DB 618 GTTTACATAGTCTTTGGGATTTTACTGTTCCCTAATTTTATTCGAAACCAATTTTACC 676

QY 662 CCAGACCATATTAATCAATTAATTAATTTGATGACAGATTGTCGAATTC 713  
 DB 677 CCAGACCATATTAATCAATTAATTAATTTGATGACAGATTGTCGAATTC 728

QY 242 TNCCTAATGATGGTGTATTGACTAATACATGATTTTTCAGAGAGAAACCATGTTA 301  
 DB 241 TTTCTTAATGATGGTGTATTGACTAATACATGATTTTTCAGAGAGAAACCATGTTA 300

QY 302 AAAATATTTTAAATTTTAAAGAAATAGCTGTGTCAGCTGATCATATTTCTTTATTT 361  
 DB 301 AAAATATTTTAAATTTTAAAGAAATAGCTGTGTCAGCTGATCATATTTCTTTATTT 360

QY 362 TGAATTTGGGAAANAAATACCTGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 421  
 DB 361 TGAATTTGGGAAANAAATACCTGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTAA 420

QY 422 TTTCCACTAATTTTAAANAACTATGAGAAATGATTAATGACATGAAGGCAACACTTAA 481  
 DB 421 TTTCCACTAATTTTAAAGAAATGATTAATGACATGAAGGCAACACTTAA 480

QY 482 TTACTGGCCAGCTGTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACCTCTTAA 541  
 DB 481 TTACTGGCCAGCTGTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACCTCTTAA 539

QY 542 ATTTGAATCTTCAGAGAAATATCTTTAATATACCTTTGTAAAGCAAAACAAAGCTTTT 601  
 DB 540 ATTTGAATCTTCAGAGAAATATCTTTAATATACCTTTGTAAAGCAAAACAAAGCTTTT 599

QY 602 GTTACATAGTCTTTGGGATTTTACTGCTCTAATTTTATTCGAAATCAATTTTACC 661  
 Db 600 GTTACATAGTCTTTGGGATTTTACTGCTCTAATTTTATTCGAAATCAATTTTACC 659  
 QY 662 CCAGACCATTAATTAACATTAATTAATTTGTTGACAGCTGTTGGCAATTC 713  
 Db 660 CCAGACCATTAATTAACATTAATTAATTTGTTGACAGCTGTTGACATTAATTC 711  
 RESULT 4  
 A1826287 770 bp mRNA linear EST 21-DEC-1999  
 LOCUS WK33f07.x1 NCI CGAP Pr22 Homo sapiens cDNA clone IMAGE:2417221.3  
 DEFINITION similar to gp:S52456 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR  
 (HMAN); contains element MER28 repetitive element ;, mRNA sequence.  
 ACCESSION A1826287  
 VERSION A1826287.1 GI:5446958  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 770)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapdb-remail.nih.gov](mailto:cgapdb-remail.nih.gov)  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
[www-bio.illn.gov/bbrp/image/image.html](http://www-bio.illn.gov/bbrp/image/image.html)  
 Insert Length: 1454 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 404.  
 Location/Qualifiers  
 1..770  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2417221"  
 /sex="male"  
 /tissue\_type="normal prostate"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI-CGAP Pr22"  
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker; 1st strand cDNA was prepared  
 from normal prostate bulk tissue, and was then primed with  
 a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT73 vector. Library is normalized, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 271 a 111 c 99 g 285 t 4 others  
 ORIGIN  
 Query Match 93.9%; Score 670.4; DB: 9; Length 770;  
 Best Local Similarity 97.8%; Pred. No. 1,46-69;  
 Matches 696; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

QY 2 GCCAGAAAAGTCTTTTAAATCTTCTTAAACATCTCTCAAGCATTTATTCCT 61  
 Db 25 GCCAGAAAAGTCTTTTAAATCTTCTTAAACATCTCTCAAGCATTTATTCCT 84  
 QY 62 ATATCTAGTGAATTTAAGAAATACATTAATTAAGAAATACATTAAGATTA 121  
 Db 85 ATATCTAGTGAATTTAAGAAATACATTAATTAAGAAATACATTAAGATTA 144

QY 122 TGCAGATTAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 181  
 Db 145 TGCAGATTAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 204  
 QY 182 AATGAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 241  
 Db 205 AATGAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 264  
 QY 242 TNCCTTAATGATGCTTTTATGACTAATTAATTAATTAATTAATTAATTAATTA 301  
 Db 265 TNCCTTAATGATGCTTTTATGACTAATTAATTAATTAATTAATTAATTAATTAATTA 324  
 QY 302 AAAATATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 361  
 Db 325 AAAATATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 384  
 QY 362 TGATTTGGGAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 421  
 Db 385 TGATTTGGGAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 444  
 QY 422 TCTCAGTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAA 481  
 Db 445 TCTCAGTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAA 504  
 QY 482 TTAATGGCCAGCTGTTGGCAATGTTGTTCTTACTTAATTTCCCAAGGAAATCTTTAA 541  
 Db 505 TTAATGGCCAGCTGTTGGCAATGTTGTTCTTACTTAATTTCCCAAGGAAATCTTTAA 563  
 QY 542 ATGAATCTTCAGAGAAATATCTTAATTAATTAATTAATTAATTAATTAATTAATTTT 601  
 Db 564 ATGAATCTTCAGAGAAATATCTTAATTAATTAATTAATTAATTAATTAATTAATTTT 623  
 QY 602 GTTACATAGTCTTTGGGATTTTACTGCTCTAATTTTATTCGAAATCAATTTTACC 661  
 Db 624 GTTACATAGTCTTTGGGATTTTACTGCTCTAATTTTATTCGAAATCAATTTTACC 682  
 QY 662 CCAGACCATTAATTAACATTAATTAATTTGTTGACAGCTGTTGGCAATTC 713  
 Db 683 CCAGACCATTAATTAACATTAATTAATTTGTTGACAGCTGTTGACATTTTC 734

RESULT 5  
 BM985376 756 bp mRNA linear EST 20-FEB-2003  
 LOCUS UI-CF-ECl-acg-p-23-0-UI.s1 UI-CF-ECl Homo sapiens cDNA clone  
 DEFINITION UI-CF-ECl-acg-p-23-0-UI 3', mRNA sequence.  
 ACCESSION BM985376  
 VERSION BM985376.1 GI:19611803  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 756)  
 Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 JOURNAL 97044477  
 MEDLINE 8889548  
 PUBMED 8889548  
 COMMENT Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: [paul-mccray@uiowa.edu](mailto:paul-mccray@uiowa.edu)  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research



Genetics (www.reagen.com) or from Open Biosystems (www.openbiosystems.com). The following repetitive elements were found in this cDNA sequence: 316-342, >AT rich#low\_complexity (matched complement) Seq primer: M13 FORWARD POLYA=Yes.

# FEATURES

source  
Location/Qualifiers  
1..756  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EC1-acg-p-23-0-UI"  
/tissue\_type="Lung"  
/dev\_stage="Adult and Fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_1lb="UI-CF-EC1"  
/note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) with a modified polylinker; Site 1: Scorer I; Site 2: Not I; UI-CF-EC1 is a normalized cDNA library containing the following tissue(s): Normal lung from adult and from fetal day 64, day 87, week 19 and week 42. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag (dT18 tail). The sequence tag for this library is AAGGCTTAC  
TAG\_LIB=UI-CF-EC1  
TAG\_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371 and 380-383  
TAG\_SEQ=AAGGCTTAC"

BASE COUNT 268 a 107 c 97 g 283 t 1 others

ORIGIN

Query Match 93.8%; Score 669.6; DB 12; Length 756;  
Best Local Similarity 97.9%; Pred. No. 1.8e-69;  
Matches 697; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

2 GCCAGAAAAGTTTAACTTCTTAATCACTCTCCAGACGTTATTTATCCCT 61  
19 GCCAGAAAAGTTTAACTTCTTAATCACTCTCCAGACGTTATTTATCCCT 78  
62 ATATCTCACTGAATTTAGAAATACATAGTATTAGAAAACTAGAAAAAGATPAA 121  
79 ATATCTCACTGAATTTAGAAATACATAGTATTAGAAAACTAGAAAAAGATPAA 138  
122 TGCAGATTAATTAACCTTACATGAAAAAGAAAATTATTAACAAGAGCTGAGACGTTATA 181  
139 TGCAGATTAATTAACCTTACATGAAAAAGAAAATTATTAACAAGAGCTGAGACGTTATA 198  
182 AATTGAATGAGATTATTAATTTGAAAACTGCATCTGAAAGCAACTTATTTGTTCAATTA 241  
199 AATTGAATGAGATTATTAATTTGAAAACTGCATCTGAAAGCAACTTATTTGTTCAATTA 258  
242 TTTCTTAATGATGCTTTTATGATCAATACCTGATTTTCAAGAGGAAACCATGTTA 301  
259 TTTCTTAATGATGCTTTTATGATCAATACCTGATTTTCAAGAGGAAACCATGTTA 318  
302 AAAAATATTTTATTTTAAAAAAGCCTGTGTCAGCTGATCATATTTCTTTATTT 361  
319 AAAAATATTTTATTTTAAAAAAGCCTGTGTCAGCTGATCATATTTCTTTATTT 378  
362 TGATTTGGGAANAAATACCTGTTCTGATACAGAAATGCAAAATTTTGGATTTTAA 421  
379 TGATTTGGGAANAAATACCTGTTCTGATACAGAAATGCAAAATTTTGGATTTTAA 438  
422 TCTCACTAATTTTAAANAACTATTGAGAAATGATTAATGACATGAAGTCAACACTAA 481  
439 TCTCACTAATTTTAAANAACTATTGAGAAATGATTAATGACATGAAGTCAACACTAA 498

QY 482 TTACTGGCCAGCTGTTGGCATGTGTTCTTACTAGTTCTCCAGGAAAGAACTTTAA 541  
DB 499 TTACTGGCCAGCTGTTGGCATGTGTTCTTACTAGTTCTCCAGGAAAGAACTTTAA 557  
QY 542 ATGAATCTTCACAGAGATATATCTTAAATATATCTTTGTAAGCAAAACAAAGCTTTT 601  
DB 558 ACTGAATCTTCACAGAGATATATCTTAAATATATCTTTGTAAGCAAAACAAAGCTTTT 617  
QY 602 GTTTTACATGCTTTCTTTGGATTTTACTGTTCTTAATTTATTTCTGAAGCAATTTTACC 661  
DB 618 GTTTTACATGCTTTCTTTGGATTTTACTGTTCTTAATTTATTTCTGAAGCAATTTTACC 676  
QY 662 CCAGACCATATATACATATTAATCTTTGTTGCAAGCTGTTCCAAATTC 713  
DB 677 CCAGACCATATATACATATTAATCTTTGTTGATGACAGTTGATGCAATTC 728

RESULT 6  
LOCUS B0045161 738 bp mRNA linear EST 21-FEB-2003  
DEFINITION UI-CF-EN1-aei-a-17-0-UI.81 UI-CF-EN1 Homo sapiens cDNA clone  
ACCESSION B0045161  
VERSION B0045161.1 GI:19796248  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 738)  
Normalization and Subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: McCray, PB

REFERENCE  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and Subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548

COMMENT  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul.mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com) or from Open Biosystems (www.openbiosystems.com).  
The following repetitive elements were found in this cDNA sequence: 316-342, >AT rich#low\_complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

source  
Location/Qualifiers  
1..738  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_1lb="UI-CF-EN1"  
/note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) with a modified polylinker; Site 1: Scorer I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an

oligo-dt primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGCT.

TAG LIB=UI-CF-EN1  
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGCT"

BASE COUNT 260 a 102 c 97 g 278 t 1 others  
ORIGIN

Query Match 93.6%; Score 668.4; DB 12; Length 738;  
Best Local Similarity 98.3%; Pred. No. 2.5e-69;  
Matches 694; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

QY 2 GCCAGAAAAAGTTTAAATTTCTATTAACATTTCTCAAGACATTAATTTATCCT 61  
DB 19 GCCAGAAAAAGTTTAAATTTCTATTAACATTTCTCAAGACATTAATTTATCCT 78  
QY 62 ATATCTCACTGAATTTAAGAAATACATAGTATTAAGAAAACTAGAAAAAGATATA 121  
DB 79 ATATCTCACTGAATTTAAGAAATACATAGTATTAAGAAAACTAGAAAAAGATATA 138  
QY 122 TGCAGATTAATTAACCTTACATGAAGAAAAATTAACAAAGACCTGAGACGTTATA 181  
DB 139 TGCAGATTAATTAACCTTACATGAAGAAAAATTAACAAAGACCTGAGACGTTATA 198  
QY 182 AATTGAAATGAGATTAATTAATTTGAAAACTGATCGAAGCAAACTTATTTGTCATTA 241  
DB 199 AATTGAAATGAGATTAATTAATTTGAAAACTGATCGAAGCAAACTTATTTGTCATTA 258  
QY 242 TNCCTAATGATGGTGTATTAAGCACTAATGATGATTTTCAAGAGGAAACCATGTTA 301  
DB 259 TTCTTAATGATGGTGTATTAAGCACTAATGATGATTTTCAAGAGGAAACCATGTTA 318  
QY 302 AAAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 361  
DB 319 AAAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 378  
QY 362 TGATTTGGAGAAAAATCTGTTCTGATGATGATGATGATGATGATGATGATGATGATG 421  
DB 379 TGATTTGGAGAAAAATCTGTTCTGATGATGATGATGATGATGATGATGATGATGATG 438  
QY 422 TCTCATAATTTTAAANAATTAAGAAATGATTAATGACATGATGATGATGATGATGATG 481  
DB 439 TCTCATAATTTTAAANAATTAAGAAATGATTAATGACATGATGATGATGATGATGATG 498  
QY 482 TTACTGGCGAGCTTTGGCATTTGTTCTTACTAGTTCTCCCAAGGAAAACTCTTAA 541  
DB 499 TTACTGGCGAGCTTTGGCATTTGTTCTTACTAGTTCTCCCAAGGAAAACTCTTAA 557  
QY 542 AATTGAATCTTCAAGAGAAATTAATCTTAAATATATCTTTGTAAGAAACAAAGCTTTT 601  
DB 558 ACTGAATCTTCAAGAGAAATTAATCTTAAATATATCTTTGTAAGAAACAAAGCTTTT 617  
QY 602 GTTTACATAGTTCTTTGGATTTTACTGTTCTTAAATTTTATTTGTAAGCTCAATTTTACC 661  
DB 618 GTTTACATAGTTCTTTGGATTTTACTGTTCTTAAATTTTATTTGTAAGCTCAATTTTACC 676  
QY 662 CCAGACCATATTTACCATATTAATCTTTGTTGACAGCTTTGTTGC 707  
DB 677 CCAGACCATATTTACCATATTAATCTTTGTTGACAGCTTTGTTGC 722

RESULT 7  
LOCUS AM052045 786 bp mRNA linear EST 20-SEP-1999  
DEFINITION w25e05.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2544704 3' similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR (HUMAN); mRNA sequence.

ACCESSION AM052045  
VERSION AM052045.1 GI:5914404  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: [www-bio.illn.gov/bdip/image/image.html](http://www-bio.illn.gov/bdip/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 458.  
Location/Qualifiers  
1. 786  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2544704"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Kid11"  
/note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid1 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Felicia Bonaldo."

BASE COUNT 290 a 116 c 109 g 270 t 1 others  
ORIGIN

Query Match 93.4%; Score 667; DB 9; Length 786;  
Best Local Similarity 97.6%; Pred. No. 3.5e-69;  
Matches 695; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 2 GCCAGAAAAAGTTTAAATTTCTATTAACATTTCTCAAGACATTAATTTATCCT 61  
DB 2 GCCAGAAAAAGTTTAAATTTCTATTAACATTTCTCAAGACATTAATTTATCCT 61  
QY 62 ATATCTCACTGAATTTAAGAAATACATTAATTAAGAAAACTAGAAAAAGATATA 121  
DB 62 ATATCTCACTGAATTTAAGAAATACATTAATTAAGAAAACTAGAAAAAGATATA 121  
QY 182 ATATCTCACTGAATTTAAGAAATACATTAATTAAGAAAACTAGAAAAAGATATA 241  
DB 182 ATATCTCACTGAATTTAAGAAATACATTAATTAAGAAAACTAGAAAAAGATATA 241  
QY 242 TNCCTAATGATGGTGTATTAAGCACTAATGATGATTTTCAAGAGGAAACCATGTTA 301  
DB 242 TNCCTAATGATGGTGTATTAAGCACTAATGATGATTTTCAAGAGGAAACCATGTTA 301  
QY 302 AAAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 361  
DB 302 AAAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 361

```

Oy 362 TGATTTGGGAAATACTGTTCTGTAGATGAAATGCAAAATTTTATGATTTTAA 421
Db 362 TGATTTGGGAAATACTGTTCTGTAGATGAAATGCAAAATTTTATGATTTTAA 421
Oy 422 TCTCACTAATTTTANMACTATTGAGAAATGATTATGACATGAGTGCACAACTTAA 481
Db 422 TCTCACTAATTTTANMACTATTGAGAAATGATTATGACATGAGTGCACAACTTAA 481
Oy 482 TTACTGGCCAGCTGTGGCATTTGTCTTCTTACTAGTTCCTCCAGGAAAACCTTTAA 541
Db 482 TTACTGGCCAGCTGTGGCATTTGTCTTCTTACTAGTTCCTCCAGGAAAACCTTTAA 540
Oy 542 ATTGAATCTCAGCAGAAATATCCTTAATATPACTTTGTAAGCAAAACAAAGCTTTT 601
Db 542 ACTGAATCTCAGCAGAAATATCCTTAATATPACTTTGTAAGCAAAACAAAGCTTTT 600
Oy 602 GTTTACATAGTTCCTTTGGATTTTACTGTTCTTAATTTTATTTGTAACATTTTACC 661
Db 602 GTTTACATAGTTCCTTTGGATTTTACTGTTCTTAATTTTATTTGTAACATTTTACC 659
Oy 662 CCAGACCAATATTTACATATTTACTTTGTTGTCACAGTTGTTGCCAATTC 713
Db 662 CCAGACCAATATTTACATATTTACTTTGTTGTAATGACAGTTGTAATTC 711

```

```

RESULT 8      751 bp      mRNA      linear      EST 26-MAR-2002
B0008197      UI-H-ED1-bfcm-1-10-0-UI-s1 NCI CGAP_ED1 Homo sapiens cDNA clone
LOCUS          IMAGE:5836866 3', mRNA sequence.
DEFINITION
ACCESSION     B0008197
VERSION       B0008197.1 GI:19733097
KEYWORDS
SOURCE        Homo sapiens
ORGANISM      Homo sapiens (human)
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 751)
AUTHORS      NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL      Tumor Gene Index
COMMENT       Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/ILM at: http://image.llnl.gov
sequence: 319-345, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLA=yes.

```

# FEATURES source

```

location/Qualifiers
1..751
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5836866"
/tissue="IMAG:5836866"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_1ib="NCI CGAP ED1"
/clone_2ib="NCI CGAP ED1"
/clone_3ib="NCI CGAP ED1"
/clone_4ib="NCI CGAP ED1"
/clone_5ib="NCI CGAP ED1"
/clone_6ib="NCI CGAP ED1"
/clone_7ib="NCI CGAP ED1"
/clone_8ib="NCI CGAP ED1"
/clone_9ib="NCI CGAP ED1"
/clone_10ib="NCI CGAP ED1"
/clone_11ib="NCI CGAP ED1"
/clone_12ib="NCI CGAP ED1"
/clone_13ib="NCI CGAP ED1"
/clone_14ib="NCI CGAP ED1"
/clone_15ib="NCI CGAP ED1"
/clone_16ib="NCI CGAP ED1"
/clone_17ib="NCI CGAP ED1"
/clone_18ib="NCI CGAP ED1"
/clone_19ib="NCI CGAP ED1"
/clone_20ib="NCI CGAP ED1"
/clone_21ib="NCI CGAP ED1"
/clone_22ib="NCI CGAP ED1"
/clone_23ib="NCI CGAP ED1"
/clone_24ib="NCI CGAP ED1"
/clone_25ib="NCI CGAP ED1"
/clone_26ib="NCI CGAP ED1"
/clone_27ib="NCI CGAP ED1"
/clone_28ib="NCI CGAP ED1"
/clone_29ib="NCI CGAP ED1"
/clone_30ib="NCI CGAP ED1"
/clone_31ib="NCI CGAP ED1"
/clone_32ib="NCI CGAP ED1"
/clone_33ib="NCI CGAP ED1"
/clone_34ib="NCI CGAP ED1"
/clone_35ib="NCI CGAP ED1"
/clone_36ib="NCI CGAP ED1"
/clone_37ib="NCI CGAP ED1"
/clone_38ib="NCI CGAP ED1"
/clone_39ib="NCI CGAP ED1"
/clone_40ib="NCI CGAP ED1"
/clone_41ib="NCI CGAP ED1"
/clone_42ib="NCI CGAP ED1"
/clone_43ib="NCI CGAP ED1"
/clone_44ib="NCI CGAP ED1"
/clone_45ib="NCI CGAP ED1"
/clone_46ib="NCI CGAP ED1"
/clone_47ib="NCI CGAP ED1"
/clone_48ib="NCI CGAP ED1"
/clone_49ib="NCI CGAP ED1"
/clone_50ib="NCI CGAP ED1"
/clone_51ib="NCI CGAP ED1"
/clone_52ib="NCI CGAP ED1"
/clone_53ib="NCI CGAP ED1"
/clone_54ib="NCI CGAP ED1"
/clone_55ib="NCI CGAP ED1"
/clone_56ib="NCI CGAP ED1"
/clone_57ib="NCI CGAP ED1"
/clone_58ib="NCI CGAP ED1"
/clone_59ib="NCI CGAP ED1"
/clone_60ib="NCI CGAP ED1"
/clone_61ib="NCI CGAP ED1"
/clone_62ib="NCI CGAP ED1"
/clone_63ib="NCI CGAP ED1"
/clone_64ib="NCI CGAP ED1"
/clone_65ib="NCI CGAP ED1"
/clone_66ib="NCI CGAP ED1"
/clone_67ib="NCI CGAP ED1"
/clone_68ib="NCI CGAP ED1"
/clone_69ib="NCI CGAP ED1"
/clone_70ib="NCI CGAP ED1"
/clone_71ib="NCI CGAP ED1"
/clone_72ib="NCI CGAP ED1"
/clone_73ib="NCI CGAP ED1"
/clone_74ib="NCI CGAP ED1"
/clone_75ib="NCI CGAP ED1"
/clone_76ib="NCI CGAP ED1"
/clone_77ib="NCI CGAP ED1"
/clone_78ib="NCI CGAP ED1"
/clone_79ib="NCI CGAP ED1"
/clone_80ib="NCI CGAP ED1"
/clone_81ib="NCI CGAP ED1"
/clone_82ib="NCI CGAP ED1"
/clone_83ib="NCI CGAP ED1"
/clone_84ib="NCI CGAP ED1"
/clone_85ib="NCI CGAP ED1"
/clone_86ib="NCI CGAP ED1"
/clone_87ib="NCI CGAP ED1"
/clone_88ib="NCI CGAP ED1"
/clone_89ib="NCI CGAP ED1"
/clone_90ib="NCI CGAP ED1"
/clone_91ib="NCI CGAP ED1"
/clone_92ib="NCI CGAP ED1"
/clone_93ib="NCI CGAP ED1"
/clone_94ib="NCI CGAP ED1"
/clone_95ib="NCI CGAP ED1"
/clone_96ib="NCI CGAP ED1"
/clone_97ib="NCI CGAP ED1"
/clone_98ib="NCI CGAP ED1"
/clone_99ib="NCI CGAP ED1"
/clone_100ib="NCI CGAP ED1"

```

cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (G)18 tail. The sequence tag for this library is GCTCAAGCT. TAG LIB=UI-H-ED1 TAG TISSUE=chondrosarcoma TAG\_SEQ=CGTCAAGCT

BASE COUNT 266 a 104 c 98 g 282 t 1 others

Query Match 93.1%; Score 664.4; DB 12; Length 751; Best Local Similarity 97.0%; Pred. No. 7.2e-69; Matches 685; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

```

Oy 2 GCCAGAAAAGTTATTTTATTTTCTATTAACATTTCTTCAAGCAATTTTATCT 61
Db 22 GCCAGAAAAGTTATTTTATTTTCTATTAACATTTCTTCAAGCAATTTTATCT 81
Oy 62 ATATCTCAGTAATTTTAAAGAAATTAACATTTAGTAATTAAGAAAGTAA 121
Db 82 ATATCTCAGTAATTTTAAAGAAATTAACATTTAGTAATTAAGAAAGTAA 141
Oy 122 TGCAGATTAATTAACATTAACATTAAGAAAGTAAAGTAAAGTAAAGTAA 181
Db 142 TGCAGATTAATTAACATTAACATTAAGAAAGTAAAGTAAAGTAAAGTAA 201
Oy 182 AATGAATGAATTAATTAATTAAGAAAGTAAAGTAAAGTAAAGTAAAGTAA 241
Db 202 AATGAATGAATTAATTAATTAAGAAAGTAAAGTAAAGTAAAGTAAAGTAA 261
Oy 242 TNCCTAATGAGTGTTTATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 301
Db 262 TTTCTAATGAGTGTTTATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 321
Oy 302 AAAATATTTTATTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 361
Db 322 AAAATATTTTATTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 381
Oy 362 TGATTTGGGAAATACTGTTCTGTAGATGAAATGCAAAATTTTATGATTTTAA 421
Db 382 TGATTTGGGAAATACTGTTCTGTAGATGAAATGCAAAATTTTATGATTTTAA 441
Oy 422 TCTCACTAATTTTANMACTATTGAGAAATGATTATGACATGAGTGCACAACTTAA 481
Db 442 TCTCACTAATTTTANMACTATTGAGAAATGATTATGACATGAGTGCACAACTTAA 501
Oy 482 TTACTGGCCAGCTGTGGCATTTGTCTTCTTACTAGTTCCTCCAGGAAAACCTTAA 541
Db 502 TTACTGGCCAGCTGTGGCATTTGTCTTCTTACTAGTTCCTCCAGGAAAACCTTAA 560
Oy 542 ATTGAATCTCAGCAGAAATATCCTTAATATPACTTTGTAAGCAAAACAAAGCTTTT 601
Db 561 ACTGAATCTCAGCAGAAATATCCTTAATATPACTTTGTAAGCAAAACAAAGCTTTT 620
Oy 602 GTTTACATAGTTCCTTTGGATTTTACTGTTCTTAATTTTATTTGTAACATTTTACC 661
Db 621 GTTTACATAGTTCCTTTGGATTTTACTGTTCTTAATTTTATTTGTAACATTTTACC 680
Oy 662 CCAGACCAATATTTACATATTTACTTTGTTGTCACAGTTGTTGCCAATTC 713
Db 681 CCAGACCAATATTTACATATTTACTTTGTTGTAATGACAGTTGTAATTC 716

```

```

RESULT 9      759 bp      mRNA      linear      EST 23-SEP-2002
B0619112      UI-H-FH1-bfm-1-10-0-UI-s1 NCI CGAP_FH1 Homo sapiens cDNA clone
LOCUS          UI-H-FH1-bfm-1-10-0-UI 3', mRNA sequence.
DEFINITION
ACCESSION     B0619112
VERSION       B0619112.1 GI:23285327
KEYWORDS
SOURCE        Homo sapiens (human)

```

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 759)  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: James Martin  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 The following repetitive elements were found in this CDNA  
 sequence: 316-342, >AT\_rich#Low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 source 1..759  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FH1-Dfm-1-10-0-UI"  
 /tissue\_type="Cell Line"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP FH1"  
 /note="Organ: Chondrosarcoma; Vector: pTT73-Pac (Pharmacia)  
 with a modified polylinker; Site 1: EcoR I; Site 2: Not  
 I; NCI CGAP FH1 is a normalized cDNA library obtained from  
 a cell line derived from grade I chondrosarcoma tissue.  
 The library was constructed and normalized according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pTT73-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is AGAATCCGCG. The cell line  
 was provided by Dr. James Martin from the University of  
 Iowa.  
 TAG\_LIB=UI-H-FH1  
 TAG\_TISSUE=Human Chondrosarcoma Cell Line C58 - Grade 1  
 Chondrosarcoma  
 TAG\_SEQ=AGAATCCGCG"

BASE COUNT 269 a 107 c 97 g 286 t

Query Match 93.1%; Score 664.4; DB 13; Length 759;  
 Best Local Similarity 97.0%; Pred. No. 7.1e-69;  
 Matches 685; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

2 GCCAGAAAAGTATTTTCTATTAACATCTCTCAAGCATTTATTCCT 61  
 19 GCCAGAAAAGTATTTTCTATTAACATCTCTCAAGCATTTATTCCT 78  
 62 ATATCTCACTGATTTTAAAGAAATTAATTAAGAAATTAAGAAATTA 121  
 79 ATATCTCACTGATTTTAAAGAAATTAATTAAGAAATTAAGAAATTA 138  
 122 TGCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 181  
 139 TGCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 198  
 182 AATTGAATGAGATTATTAATTAATTAATTAATTAATTAATTAATTA 241  
 199 AATTGAATGAGATTATTAATTAATTAATTAATTAATTAATTAATTA 258

242 TNCCTAATGATGATGATTTTAAAGCACTAATACCTGATTTTCAAGAAAGCAACCTGTTA 301  
 259 TTCTTAATGATGATGATTTTAAAGCACTAATACCTGATTTTCAAGAAAGCAACCTGTTA 318  
 302 AAAATATTTTATTTTAAAGAAATTAATTAAGCACTGATTTTCAAGCACTGATTTTCAAG 361  
 319 AAAATATTTTATTTTAAAGAAATTAATTAAGCACTGATTTTCAAGCACTGATTTTCAAG 378  
 352 TGATTTGGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 421  
 379 TGATTTGGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 438  
 422 TCTCACTAATTTTAAAGCACTGATTTTCAAGCACTGATTTTCAAGCACTGATTTTCAAG 481  
 439 TCTCACTAATTTTAAAGCACTGATTTTCAAGCACTGATTTTCAAGCACTGATTTTCAAG 498  
 482 TTACTGGCAGCTGTTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACTCTTAA 541  
 499 TTACTGGCAGCTGTTGGCATTTGTTTCTTACTAGTTCTCCAGAGGAAACTCTTAA 557  
 542 ATTGAATTTTCAAGCACTGATTTTCAAGCACTGATTTTCAAGCACTGATTTTCAAG 601  
 558 ACTGAATTTTCAAGCACTGATTTTCAAGCACTGATTTTCAAGCACTGATTTTCAAG 617  
 602 GTTACATAGTTCTTGGGATTTTACTGTTCTAATTTTATTTGAACCTCAATTTTACC 661  
 618 TGTTCATAGTTCTTGGGATTTTACTGTTCTAATTTTATTTGAACCTCAATTTTACC 677  
 662 CCAGACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 707  
 678 CCAGACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 723

RESULT 10  
 BM968746 721 bp mRNA linear EST 20-FEB-2003  
 LOCUS UI-CF-DU1-aam-a-03-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone  
 DEFINITION UI-CF-DU1-aam-a-03-0-UI 3', mRNA sequence.  
 ACCESSION BM968746  
 VERSION BM968746.1 GI:19586333  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 721)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: Paul.mccray@uiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics ([www.resgen.com](http://www.resgen.com)) or from Open Biosystems  
 ([www.openbiosystems.com](http://www.openbiosystems.com)).  
 The following repetitive elements were found in this CDNA  
 sequence: 315-341, >AT\_rich#Low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.  
 Location/Qualifiers  
 1..721  
 /organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aam-a-03-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone.lib="UI-CF-DUI"
/note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells. The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pRT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG.lib=UI-CF-DUI
TAG.TISSUE=Lung Epithelial Cells tissue nos 359-368
TAG_SEQ=GGCTGTAGGC"
BASE COUNT      257 a      97 c      91 g      276 t
ORIGIN
Query Match      93.0%; Score 664.2; DB 12; Length 721;
Best Local Similarity 98.0%; Pred. No. 7.7e-69;
Matches 691; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

3 CCAGAAAAGTATTTTATTTTCTTAAATCTTCTCAACCATTTATTTCTCTA 62
19 CCAGAAAAGTATTTTATTTTCTTAAATCTTCTCAACCATTTATTTCTCTA 78
63 TATCTCAGTATTTTAAAGAAATACATTAGATTAGAAAACAGAAAAGATAAT 122
79 TATCTCAGTATTTTAAAGAAATACATTAGATTAGAAAACAGAAAAGATAAT 138
123 GCAGATTAATTAACCTTACATGAAAAAGAAAATTAACAAAGAGCTGAAAGCTTATTA 182
139 GCAGATTAATTAACCTTACATGAAAAAGAAAATTAACAAAGAGCTGAAAGCTTATTA 198
183 ATTGAAAAGAGTTTATTTGAAAACCTGATCTGAAAAGCAACTTTATTTGTCATTTAT 242
199 ATTGAAAAGAGTTTATTTGAAAACCTGATCTGAAAAGCAACTTTATTTGTCATTTAT 258
243 NCTTAATGATGTTTATGACTAATACATGATTTTTCAGAGAAAGCAACATGTTTAA 302
259 TCTTAATGATGTTTATGACTAATACATGATTTTTCAGAGAAAGCAACATGTTTAA 318
303 AAATATTTTATTTTAAATAAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTT 362
319 AAATATTTTATTTTAAATAAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTT 378
422 GATTGGGAGAAAATCTGTTTCTGATGCAATGAAAATGCAAAATTTTATTTTAT 422
379 GATTGGGAGAAAATCTGTTTCTGATGCAATGAAAATGCAAAATTTTATTTTAT 438
423 CTCACATATTTTAAANAAGTATGAGAAATGTTATGATGATGAGTGCACAACTAT 482
439 CTCACATATTTTAAANAAGTATGAGAAATGTTATGATGATGAGTGCACAACTAT 498
483 TACTGGCAGCTGTGGCATTGTGTTCTTACTTATGTTCTCCCAAGGAAAACTCTTAA 542
499 TACTGGCAGCTGTGGCATTGTGTTCTTACTTATGTTCTCCCAAGGAAAACTCTTAA 557
543 TTGAATCTTCAGCAGAAATATCTTAAATATCTTTGTAAGAAAACAAAGCTTTTGG 602
558 CTGAATCTTCAGCAGAAATATCTTAAATATCTTTGTAAGAAAACAAAGCTTTTGG 617
603 TTACATATGTTCTTTGGGATTTTACGTTCTTAATTTTATTTGAAACATTTTACCC 662

```

```

Db      618 TTACATAGTTCTTT-GGATTTTACTGTTCTCTAATTTTATTCGAAACTCATTTTACCC 676
Qy      663 CAGACCAATATTTACCATTTTACTTTGTTTNGCAGCTGTTTGC 707
Db      677 CAGACCAATATTTTACCATTTTACTTTGTTTNGCAGCTGTTTGC 721

RESULT 11
LOCUS   CA424156
DEFINITION
UI-H-FEI-bdv-m-12-0-UI.s1 NCI CGAP FEI Homo sapiens cDNA clone
UI-H-FEI-bdv-m-12-0-UI 3', mRNA sequence.
ACCESSION
CA424156
VERSION
CA424156.1 GI:24766882
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE
Unpublished
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cgep@bbs-remail.nih.gov
Tissue Procurement: James Martin
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@iowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-48, >AT-rich#low_complexity (matched complement)
314-340, >AT-rich#low_complexity (matched complement)
Seq primer: M13 FORWARD
POLY=A=yes.

FEATURES
source
location/Qualifiers
1..734
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FEI-bdv-m-12-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone.lib="NCI CGAP FEI"
/note="Organ: Chondrosarcoma; Vector: pRT73-Pac (Pharmacia)
with a modified polylinker; Site_1: EcoR I; Site_2: Not
I; NCI CGAP FEI is a normalized cDNA library derived from
a pool of mRNA obtained from 3 cell lines from grade II
chondrosarcoma tissues. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pRT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GGCTGTAGGC. The cell
lines were provided by Dr James Martin from the University
of Iowa.
TAG.lib=UI-H-FEI
TAG.TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG_SEQ=GGCTGTAGGC"
BASE COUNT      261 a      101 c      94 g      277 t      1 others
ORIGIN
Query Match      92.9%; Score 663.2; DB 14; Length 734;
Best Local Similarity 98.0%; Pred. No. 1e-68;
Matches 690; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

```

```

Oy 4 CAGAAAAGTATTTTATTTCTATTAACATCTCTCAAGCATTTATTCCTAT 63
Db 19 CAGAAAAGTATTTTATTTCTATTAACATCTCTCAAGCATTTATTCCTAT 78
Oy 64 ATCTACGATTTTAAAGAAATTAACATTTGATTTGAAAACTAGAAAAAGATPAATG 123
Db 79 ATCTACGATTTTAAAGAAATTAACATTTGATTTGAAAACTAGAAAAAGATPAATG 138
Oy 124 CAGATTAATTAACCTACATGAAAAAGAAAAATTAACAAAGACCTGAAAGCTTATAA 183
Db 139 CAGATTAATTAACCTACATGAAAAAGAAAAATTAACAAAGACCTGAAAGCTTATAA 198
Oy 184 TTGAATAGATTAATTAATTTGAAAACTGATCTGAAAGCAAACTTAAATGTTCAATTATN 243
Db 199 TTGAATAGATTAATTAATTTGAAAACTGATCTGAAAGCAAACTTAAATGTTCAATTATN 258
Oy 244 CTTAATGATGTTTATTTATGCTAATTAACATGATTTTTCAGAGAAAAAGCCATGTAAA 303
Db 259 CTTAATGATGTTTATTTATGCTAATTAACATGATTTTTCAGAGAAAAAGCCATGTAAA 318
Oy 304 AATATTTTATTTTAAAAAAGCCTGTTCAAGCTCGATCATATTTCTTTATTTTG 363
Db 319 AATATTTTATTTTAAAAAAGCCTGTTCAAGCTCGATCATATTTCTTTATTTTG 378
Oy 364 ATTGGGAANAATACTGTTCTGATAGCATGAATGCAAAATTTTGAATTTTAAATC 423
Db 379 ATTGGGAANAATACTGTTCTGATAGCATGAATGCAAAATTTTGAATTTTAAATC 438
Oy 424 TCACATATTTTAAACATTTAGCAAAATTTGATTAATGACATGAGTGCACAACTTAAT 483
Db 439 TCACATATTTTAAACATTTAGCAAAATTTGATTAATGACATGAGTGCACAACTTAAT 498
Oy 484 ACTGGCCAGCTGTTGGCATGTTGTTCTTACTTACTGTTCCCAAGGAAAACTCTTAAT 543
Db 499 ACTGGCCAGCTGTTGGCATGTTGTTCTTACTTACTGTTCCCAAGGAAAACTCTTAAT 557
Oy 544 TGAATCTTCACAGATAATCTTAAATATACTTTGTAAGCAAAACAAAGCTTTTGT 603
Db 558 TGAATCTTCACAGATAATCTTAAATATACTTTGTAAGCAAAACAAAGCTTTTGT 617
Oy 604 TTACATAGTTCTTTGGGATTTTACGTTCTTAATTTTATCTGAAACTCAATTTTACCC 663
Db 618 TTACATAGTTCTTTGGGATTTTACGTTCTTAATTTTATCTGAAACTCAATTTTACCC 676
Oy 664 AGACATTAATTAACATTAACCTGTTGACAGCTGTTGC 707
Db 677 CAGACATTAATTAACATTAACCTGTTGACAGCTGTTGC 720

RESULT 12
BU627147 762 bp mRNA linear EST 23-SEP-2002
LOCUS UI-H-FG0-bct-1-12-0-UI.s1 NCI_GAP_EN1.2 Homo sapiens cDNA clone
DEFINITION UI-H-FG0-bct-1-12-0-UI 3', mRNA sequence.
ACCESSION BU627147
VERSION BU627147.1 GI:23293361
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapb-remail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained

```

from Dr. M. Bento Soares, bentso-soares@iowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-43, >AT-rich#low-complexity (matched complement)  
 309-335, >AT-rich#low-complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

## SOURCE

Location/Qualifiers  
 1..762  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FG0-bct-1-12-0-UI"  
 /tissue\_type="Enchondroma cell line"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI\_GAP\_EN1.2"  
 /note="Organ: Bone; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: Ecor I; Site 2: Not I;  
 NCI GAP EN1.2 is a cDNA library containing the following  
 tissue(s): Enchondroma cell line (2 cell lines). The  
 library was constructed according to Bonaldo, Lennon and  
 Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an Ecor I adaptor, digested with Not I, and cloned  
 directionally into pT73-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dr)18 tail. The sequence tag for this  
 library is CCGGTCATC. The cell lines was provided by Dr  
 James Martin from University of Iowa.  
 TAG\_Lib=UI-H-FG0  
 TAG\_Tissue=Enchondroma cell line (Mix of EN1 and EN2)  
 TAG\_SeQ=CGGTCATC"

```

BASE COUNT 269 a 109 c 97 g 286 t 1 others
ORIGIN
Query Match 92.6%; Score 661.4; DB 13; Length 762;
Best Local Similarity 98.3%; Pred. No. 1.6e-66;
Matches 687; Conservative 0; Mismatches 10; Indels 2; Gaps 2;
Oy 9 AAGTTATTTTAAATTTCTATTAACATCTCTCAAGCATTTATTCCTATCTC 68
Db 19 AAGTTATTTTAAATTTCTATTAACATCTCTCAAGCATTTATTCCTATCTC 78
Oy 69 ACTGAATTTTAAAGAAATTAACATTTGATTAATGAAAACTAGAAAAAGATPAATG 128
Db 79 ACTGAATTTTAAAGAAATTAACATTTGATTAATGAAAACTAGAAAAAGATPAATG 138
Oy 129 AATTAACCTACATGAAAAAGAAAAATTAACAAAGACCTGAAAGCTTATAA 188
Db 139 AATTAACCTACATGAAAAAGAAAAATTAACAAAGACCTGAAAGCTTATAA 198
Oy 189 ATGAGATTAATTAATTTGAAAACTGATCTGAAAGCAAACTTAAATGTTCAATTATN 248
Db 199 ATGAGATTAATTAATTTGAAAACTGATCTGAAAGCAAACTTAAATGTTCAATTATN 258
Oy 249 TGATGTTGTTTATGCTAATTAACATGATTTTTCAGAGAAAAAGCCATGTAAAATAT 308
Db 259 TGATGTTGTTTATGCTAATTAACATGATTTTTCAGAGAAAAAGCCATGTAAAATAT 318
Oy 309 TTTTATTTTAAAAAAGCCTGTTGTCACCTCGATCATATTTCTTTATTTGATTTG 366
Db 319 TTTTATTTTAAAAAAGCCTGTTGTCACCTCGATCATATTTCTTTATTTGATTTG 378
Oy 369 GGAANAATACTGTTCTGATAGCATGAATGCAAAATTTTGAATTTTAAATCTGACT 428
Db 379 GGAANAATACTGTTCTGATAGCATGAATGCAAAATTTTGAATTTTAAATCTGACT 438
Oy 429 AATTTTAAACATTTAGCAAAATTTGATTAATGACATGAGTGCACAACTTAATCTGG 488
Db 439 AATTTTAAACATTTAGCAAAATTTGATTAATGACATGAGTGCACAACTTAATCTGG 498

```

489 CCAGCTGTGGCATTGTGTTCTTACTAGTTCCTCCAGGAAACCTTTAATTGAT 548  
499 CCAGCTGTGGCATTGTGTTCTTACTAGTTCCTCCAGGAAACCTTTAATTGAT 557  
549 CTTGACGAGAAATATCCTTAATATACCTTTGTAAGCAAAACAAAGCTTTTGTTCACA 608  
558 CTTGACGAGAAATATCCTTAATATACCTTTGTAAGCAAAACAAAGCTTTTGTTCACA 617  
609 TAGTCTTTGGGATTTTACTGTTCTTCTAATTTTATCTGAAACTCAATTTTACCCAGACC 668  
618 TAGTCTTTGGGATTTTACTGTTCTTCTAATTTTATCTGAAACTCAATTTTACCCAGACC 676  
669 AATAATTACATATTACTTTGTTTNGACACAGTTGTTTC 707  
677 AATAATTACATATTACTTTGTTTNGACACAGTTGTTTC 715

RESULT 13  
BU933572/c 734 bp mRNA linear EST 18-OCT-2002  
LOCUS BU933572 IMAGE:6698694 5', mRNA sequence.  
DEFINITION AGENCOURT 10507105 NIH\_MGC\_127 Homo sapiens cDNA clone  
ACCESSION BU933572  
VERSION BU933572.1 GI:24122391  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 734)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: NCI  
cDNA Library Preparation: Michael Brownstein Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LNCM2985 row: f column: 06  
High quality sequence stop: 526.  
Location/Qualifiers  
1..734  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone IMAGE:6698694"  
/tissue\_type="mixed (pool of 40 RNAs)"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_1ib="NIH MGC 127"  
/note="Vector: pDNR-LIB; Site 1: SfiI (ggccatcatgccc);  
Site 2: SfiI (ggccgcctggccc); Double-stranded cDNA was  
prepared from a pool of 40 cell line polyA+ RNAs (bladder  
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -  
4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,  
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,  
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary  
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were  
used in cloning as follows:  
5'-AAGCAGTGTATCAACGAGAGTGGCATTACGCGCGG-3' and  
5'-ATTCTAGAGCGCGAGCGCGCGACATG-dT(30)NN-3'. Full-length  
enriched library was constructed using the Clontech  
Creator SMART kit and size-selected to contain the 1-2 kb  
size fraction (other fractions present in NIH\_MGC\_126 and  
NIH\_MGC\_128). Library created in the laboratory of T.  
Uedin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH\_MGC  
Library."

BASE COUNT 271 a 98 c 101 g 256 t 8 others  
ORIGIN

Query Match 92.5%; Score 660.6; DB 13; Length 734;  
Best Local Similarity 98.4%; Pred. No. 2e-68;  
Matches 685; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

2 GCCAGAAAAGTATTTTAAATTTTCTATTAACATCTTCTGAAAGCATTTATTCCT 61  
695 GCCAGAAAAGTATTTTAAATTTTCTATTAACATCTTCTGAAAGCATTTATTCCT 636  
62 ATATCTACGAAATTTTAAAGAAATTAACATAGTCTTGAAGAAATAGAAAGATTA 121  
635 ATATCTACGAAATTTTAAAGAAATTAACATAGTCTTGAAGAAATAGAAAGATTA 576  
122 TGCAGATTAATTAACATTAACATGAAGAAATTAATTAACAAAGACTGAAGCTTATA 181  
575 TGCAGATTAATTAACATTAACATGAAGAAATTAATTAACAAAGACTGAAGCTTATA 516  
182 AATTGAATGAGATTAATTAATTTGAAAACCTGCATCTGAAGCAACCTTATTTCAATTA 241  
515 AATTGAATGAGATTAATTAATTTGAAAACCTGCATCTGAAGCAACCTTATTTCAATTA 456  
242 TNCCTTAATGATGCTGTTTATGACTAATACATGATTTTTCAGAAAGAACCATGTTA 301  
455 TTCTTAATGATGCTGTTTATGACTAATACATGATTTTTCAGAAAGAACCATGTTA 396  
302 AAATATTTTATTTTAAATTAAGCTGCTGCTGATCATATTTCTTTATTT 361  
395 AAATATTTTATTTTAAATTAAGCTGCTGCTGATCATATTTCTTTATTT 336  
362 TGATTTGGGAAATTAATCTGTTTCTGATAGCATGAATGCAAAATTTTGAATTTTAA 421  
335 TGATTTGGGAAATTAATCTGTTTCTGATAGCATGAATGCAAAATTTTGAATTTTAA 276  
422 TCTACATATTTTAAATTAAGCTGTTGAAATTTGATTAATGATGAAGCAACCATTA 481  
275 TCTACATATTTTAAATTAAGCTGTTGAAATTTGATTAATGATGAAGCAACCATTA 216  
482 TTACTGGGACGCTGTGGCATGTGTTCTTACTAGTTCCTCCAGGAAACCTTTAA 541  
215 TTACTGGGACGCTGTGGCATGTGTTCTTACTAGTTCCTCCAGGAAACCTTTAA 157  
542 ATGATATCTTCAGAGATTAATCTTAAATTAATTAATTAATTAATTAATTAATTAAT 601  
156 ACTGATCTTCAGAGATTAATCTTAAATTAATTAATTAATTAATTAATTAATTAAT 97  
602 GTTACATAGTCTTTGGGATTTTACTGTTCTTAATTTTATTTTGAAGCAATTTTACC 661  
96 GTTACATAGTCTTTGGGATTTTACTGTTCTTAATTTTATTTTGAAGCAATTTTACC 38  
662 CCAGACCATATTACATATTACTTTGTTNGAC 697  
37 CCAGACCATATTACATATTACTTTGTTNGAC 2

RESULT 14  
BU619082 712 bp mRNA linear EST 23-SEP-2002  
LOCUS BU619082  
DEFINITION UI-H-FH1-bfm-f-06-0-UI.s1 NCI CGAP FH1 Homo sapiens cDNA clone  
UI-H-FH1-bfm-f-06-0-UI 3', mRNA sequence.  
ACCESSION BU619082  
VERSION BU619082.1 GI:23285297  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 712)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: James Martin



cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@iowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-32, >AT rich#low complexity (matched complement)  
 298-324, >AT rich#low complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=yes.

# FEATURES

source

Location/Qualifiers  
 1..712  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UT-H-FH1-bfm-f-06-0-UT"  
 /tissue\_type="Cell Line"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP FH1"  
 /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac (Pharmacia)  
 ) with a modified polylinker; Site 1: EcoR I; Site 2: Not  
 I; NCI CGAP FH1 is a normalized cDNA library obtained from  
 a cell line derived from grade I chondrosarcoma tissue.  
 The library was constructed and normalized according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pT7T3-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (GT)<sub>18</sub> tail. The  
 sequence tag for this library is AGAATCCGGC. The cell line  
 was provided by Dr. James Martin from the University of  
 Iowa.  
 TAG\_lib=UI-H-FH1  
 TAG\_tissue=Human Chondrosarcoma Cell Line C88 - Grade 1  
 Chondrosarcoma  
 TAG\_SEQ=AGATCCGGC"

BASE COUNT 253 a 253 g 89 g 271 t  
 ORIGIN

Query Match 91.5%; Score 653.4; DB 13; Length 712;  
 Best Local Similarity 97.0%; Pred. No. 1.4e-67;  
 Matches 674; Conservative 0; Mismatches 20; Indels 1; Gaps 1;  
 13 TTATTTTAAATTTTCATTAAACATTTCTCAAGCATTATTTATCCTATATCTCACTG 72  
 12 TTATTTTAAATTTTCATTAAACATTTCTCAAGCATTATTTATCCTATATCTCACTG 71  
 73 AATTTTAAAGAAATACATTAGTATTAGAAAACTTGAAAAAGATTAATGCAATATT 132  
 72 AATTTTAAAGAAATACATTAGTATTAGAAAACTTGAAAAAGATTAATGCAATATT 131  
 133 AAACCTTCATGAAAAAGAAATATTAACAAGAGCTGAGAACCTTTAATTAATGGA 192  
 132 AAACCTTCATGAAAAAGAAATATTAACAAGAGCTGAGAACCTTTAATTAATGGA 191  
 193 GATTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGTCATTAATCTTAATGAT 252  
 192 GATTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGTCATTAATCTTAATGAT 251  
 253 GGCTTTTATGACTAATACACTGATTTTCAAGAGAAACCCATGTTAAAAATATTTT 312  
 252 GGCTTTTATGACTAATACACTGATTTTCAAGAGAAACCCATGTTAAAAATATTTT 311  
 313 AATTTTAAATTAAGCCGTGTTCAAGCTCGATATATTTCTTTATTTTGGGAA 372  
 312 AATTTTAAATTAAGCCGTGTTCAAGCTCGATATATTTCTTTATTTTGGGAA 371  
 373 NAAATAATCTGTTCTGATAGCAAAATGCAAAATTTTGAATTTTAACTCACTAAT 432

Db 372 GAAATACTGTTCTGATAGCAAAATGCAAAATTTTATGATTTTAACTCACTAAT 431  
 Qy 433 TTAAACTATTTGAAATTTGATTAATGACATGAGTGACACACACATTAATCTGGCCAG 492  
 Db 432 TTAAAGAACTATTTGAAATTTGATTAATGACATGAGTGACACACACATTAATCTGGCCAG 491  
 Qy 493 CTGTTGGCATGTGTTTCTTCTTCTAGTTCCTCCAAAGGAAACCTTAATGTAATCTTC 552  
 Db 492 CTGTTGGCATGTGTTTCTTCTTCTAGTTCCTCCAAAGGAAACCTTAATGTAATCTTC 550  
 Qy 553 AGCAGAAATATCTTAAATATATCTTGTGAAGCAAAAGCAAAAGCTTTTGTGTACATAGT 612  
 Db 551 AGCAGAAATATCTTAAATATATCTTGTGAAGCAAAAGCAAAAGCTTTTGTGTACATAG 610  
 Qy 613 TCTTTGGCATTTTCTGTTCTCTAATTTTATTTGTAATCTCAATTTTACCCAGACATAA 672  
 Db 611 TCTTTGGCATTTTCTGTTCTCTAATTTTATTTGTAATCTCAATTTTACCCAGACATAA 670  
 Qy 673 TTACCATTTAATCTTGTGTTGACAGAGTTGTTGC 707  
 Db 671 TTACCATTTAATCTTGTGTTGACAGAGTTGTTATGC 705

RESULT 15  
 LOCUS BQ009270 750 bp mRNA linear EST 26-MAR-2002  
 DEFINITION UT-H-ED1-axx-n-02-0-UT.s1 NCI CGAP ED1 Homo sapiens cDNA clone  
 ACCESSION BQ009270  
 VERSION BQ009270  
 KEYWORDS BQ009270.1 GI:19734171  
 SOURCE EST.  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 750)  
 NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-rc@mail.nih.gov  
 Tissue Procurement: Dr. Jose Mercuende.  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LINL at: http://image.lnll.gov  
 The following repetitive elements were found in this cDNA  
 sequence: 316-342, >AT rich#low complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=yes.

JOURNAL  
 COMMENT

## FEATURES

source

Location/Qualifiers  
 1..750  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5835073"  
 /tissue\_type="Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP ED1"  
 /note="Organ: Left Pubic Bone; Vector: pT7T3-Pac  
 (Pharmacia) with a modified polylinker; Site 1: EcoR I;  
 Site 2: Not I; NCI CGAP ED1 is a normalized cDNA library  
 containing the following tissue(s): Chondrosarcoma cell  
 line C88. The library was constructed according to Bonaldo  
 and Soares, Genome Research, 6:791-806, 1996.  
 First strand cDNA synthesis was primed with an oligo-dT  
 primer containing a Not I site. Double stranded cDNA was  
 ligated to an EcoR I adaptor, digested with Not I, and  
 cloned directionally into pT7T3-Pac vector. The  
 oligonucleotide used to prime the synthesis of



first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dfr)18 tail. The sequence tag for this library is GCTCAAGGCT.

TAG\_Lib=UT-H-ED1  
TAG\_TISSUE=chondrosarcoma  
TAG\_SEQ=CGTCAAGGCT"

BASE COUNT 266 a 105 c 96 g 281 t 2 others  
ORIGIN

Query Match 91.2%; Score 651.2; DB 12; Length 750;  
Best Local Similarity 97.1%; Pred. No. 2.5e-67;  
Matches 692; Conservative 0; Mismatches 18; Indels 3; Gaps 3;

```
QY 2 GCCAGAAAAGTTATTTTCTATTAACATTCCTCAAGCATTTTATTCCT 61
Db 19 GCCAGAAAAGTTATTTTCTATTAACATTCCTCAAGCATTTTATTCCT 78
QY 62 ATATCTCACTGAATTTTAAAGAAATACATTAAGTAAAGTAAAGTAA 121
Db 79 ATATCTCACTGAATTTTAAAGAAATACATTAAGTAAAGTAAAGTAA 138
QY 122 TGCAGATTAATTAACCTTACATGAAAAAGAAAAATTATACAAAGACTGAGAACCTTATA 181
Db 139 TGCAGATTAATTAACCTTACATGAAAAAGAAAAATTATACAAAGACTGAGAACCTTATA 198
QY 182 AATTGAATGAGATTATTAATTTGAAAACCTGATCTGAAAGCAACTTATTTGTCATTA 241
Db 199 AATTGAATGAGATTATTAATTTGAAAACCTGATCTGAAAGCAACTTATTTGTCATTA 258
QY 242 TNCCTAATGATGCTTTTATGATTAATACATGATTTTCAAGAGAAACCATGTTA 301
Db 259 TTCTTAATGATGCTTTTATGATTAATACATGATTTTCAAGAGAAACCATGTTA 318
QY 302 AAAATATTTTATTTTAAAAAATAAGCTGTGTCAGCTGATCATATTTCTTTATTT 361
Db 319 AAAATATTTTATTTTAAAAAATAAGCTGTGTCAGCTGATCATATTTCTTTATTT 378
QY 362 TGATTTGGGAAATAATCTGTTCTGATAGCATGAAATGCAAAATTTTATTTTAA 421
Db 379 TGATTTGGGAAATAATCTGTTCTGATAGCATGAAATGCAAAATTTTATTTTAA 438
QY 422 TCTCACTAATTTTAAANAACTATTGAGAAATGATTAATGACATGAGTCAACACTAA 481
Db 439 TCTCACTAATTTTAAAGAACTATTGAGAAATGATTAATGACATGAGTCAACACTAA 498
QY 482 TTACTGGCCAGCTGTGGCATGTGTCTTACTTAAGTTCTCCCAAGGAAAACTTTAA 541
Db 499 TTACTGGCCAGCTGTGGCATGTGTCTTACTTAAGTTCTCCCAAGGAAAACTTTAA 557
QY 542 ATTGAATCTTACGAGAAATATCTTAAATATACTTTGTAAGCAAAACAAAGCTTTT 601
Db 558 ACTGAATCTTACGAGAAATATCTTAAATATACTTTGTAAGCAAAACAAAGCTTTT 617
QY 602 GTTACATAGTCTTTGGGATTTTACTGTTCTTATTTTATTTGAAACTCA-TTTTAC 660
Db 618 GTTACATAGTCTTT-GGATTTTACTGTTCTTATTTTANCTGAAACTCAATTTTAC 676
QY 661 CCCAGACCATTAATTAACATTAATTTGTTTNGCAGAGTTGTTGCCAATTC 713
Db 677 CCCAGACCATTAATTAACATTAATTTGTTTNGCAGAGTTGTTGCCAATTC 729
```

Search completed: November 27, 2003, 12:28:11  
Job time : 1930.36 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:01:24 : Search time 234.619 Seconds  
(without alignments)  
8215.019 Million cell updates/sec

Title: US-09-835-992A-19

Perfect score: 714  
Sequence: 1 cgcacgaataaaagttatttca.....cacagttgttcacattca 714

Scoring table: IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 510512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

1:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*
25:	/SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710	99.4	714	20	AAAX40094
2	674.4	94.5	2663	22	AAAX57501
3	611.8	85.7	687	24	AAAX40095
4	478.2	67.0	486	24	ABV87267
5	436.2	61.1	843	20	AAAX40097
6	313.2	43.9	351	16	AAT19975
7	258.4	36.2	262	22	AAH57293
8	165.6	23.2	173	25	ABX83093

9	126.2	17.7	397	25	ABX40797
10	107.4	15.0	263	15	AAO76489
11	77.8	10.8	8056	25	ABZ10246
12	76.8	10.8	397	25	ABX47851
13	75.2	10.5	449	25	ABX48095
14	73.4	10.3	2152	24	AB199762
15	71	9.9	8056	25	ABZ10246
16	70.2	9.8	8056	25	ABZ10100
17	67.6	9.5	883	22	AA15210
18	67.6	9.5	50000	24	ABU56201
19	67.4	9.4	4985	24	ABO75107
20	66.4	9.3	5852	12	AAQ11710
21	65.2	9.1	7676	24	ABL70409
22	65.2	9.1	7676	24	ABL4598
23	64.6	9.0	6106	22	AA564439
24	64.6	9.0	6106	24	ABK40011
25	64.6	9.0	6106	24	ABL3472
26	64.6	9.0	13131	24	ABL92249
27	64.2	9.0	8056	25	ABZ10100
28	63.6	8.9	6944	24	ABK40026
29	62.8	8.8	1501	25	ABZ10188
30	62.8	8.8	50000	24	ABL56201
31	62.2	8.7	291	25	ABX48529
32	62.2	8.7	10048	24	ABL70313
33	62.2	8.7	10048	24	AA561251
34	61.8	8.7	5962	24	ABL33287
35	61.4	8.6	17721	24	ABL33729
36	61.2	8.6	20674	21	AAC58017
37	60.8	8.5	424	25	ABK46053
38	60.6	8.5	10467	24	ABK28453
39	59.6	8.3	6169	22	AA546370
40	59.6	8.3	6169	24	ABN80097
41	59.6	8.3	6912	24	ABK28371
42	59.6	8.3	7631	24	ABL32861
43	59.4	8.3	12138	24	ABL33943
44	59.4	8.3	12138	24	ABK28336
45	59.2	8.3	5815	24	ABK40024

## ALIGNMENTS

RESULT 1  
ID AAX40094 standard; DNA, 714 BP.  
XX AAX40094;  
XX  
XX  
DT 02-JUL-1999 (first entry)  
XX  
XX Gastric cancer associated gene.  
DE  
XX Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN MO9904265-A2.  
XX  
XX PD 28-JAN-1999.  
XX  
XX PF 15-JUL-1998; 98WO-US14679.  
XX  
XX PR 22-JUN-1998; 98US-0102322.  
PR 17-JUL-1997; 97US-0896164.  
PR 10-OCT-1997; 97US-0061599.  
PR 10-OCT-1997; 97US-0061765.  
PR 10-OCT-1997; 97US-0948705.  
PR 11-OCT-1997; 97GB-0021697.  
XX  
XX PA (LUDW-) LUDWIG INST CANCER RES.  
XX

Bovine EST associa  
Human genome fragm  
Haematopoietic cel  
Bovine EST associa  
Bovine EST associa  
Bovine EST associa  
Mouse ischaemic co  
Haematopoietic cel  
Haematopoietic cel  
Human breast cance  
Human genome fragm  
Anopheles gambiae  
Dictyostellium plas  
Chemically treated  
Human metastasis a  
Tumour suppressor  
Human chemically p  
Human immune syste  
Chemically treated  
Haematopoietic cel  
Human DNA for atag  
Haematopoietic cel  
AmePV genome fragm  
Bovine EST associa  
Chemically treated  
Human gene regulat  
Human immune syste  
Arachidonic acid m  
Bovine EST associa  
DNA transcription  
Tumour suppressor  
Human chemically m  
DNA transcription  
Human immune syste  
Human immune syste  
DNA transcription  
Human chemically p





Db 540 CNGA7A-A-CTTAATATACCTTGTAGCCAAACAAA-CTTTTGTTCATAGTTC 597  
Qy 615 TTGGGATTTTACTGTTCTCTATTTTATTCGAACCTCAATTTTACCCAGCCATTAAT 674  
Db 598 TTT-GGATTTTAACTGTTCTCTATTTTATTCGAACCTCAATTTTCCCGACCAATTAAT 656  
Qy 675 ACCATA-TTAACTTGTNTGACAGTGTGTT 704  
Db 657 ACCCTATTAACTTGTGTATGACAGTGTGTT 687

RESULT 4  
ABV87267  
ID ABV87267 standard; cDNA; 486 BP.  
AC ABV87267;  
XX  
XX  
XX 13-DEC-2002 (first entry)  
XX  
XX Human colon cancer related cDNA SEQ ID NO 578.  
XX  
XX Human; colon; cancer; cytoskeletal; tumour; gene therapy; vaccine; gene;  
XX ss.  
XX Homo sapiens.  
XX WO200258534-A2.  
XX  
XX 01-AUG-2002.  
XX  
XX 19-NOV-2001; 2001WO-US43704.  
XX  
XX 20-NOV-2000; 2000US-252222P.  
XX 06-FEB-2001; 2001US-267011P.  
XX 28-MAR-2001; 2001US-279670P.  
XX 10-JUL-2001; 2001US-304037P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;  
XX WPI; 2002-608400/65.  
XX  
XX New isolated tumor colon polynucleotide and polypeptide, useful for the  
XX diagnosis, prevention and/or treatment of cancer, in particular colon  
XX cancer -  
XX  
XX Claim 1; SEQ ID NO 578; 266bp + Sequence Listing; English.  
XX  
XX The invention relates to a human colon tumour expressed polynucleotide  
XX (I) encoding a polypeptide (II), ABP6791-ABP6796) comprising: (i) any of  
XX 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)  
XX complements of (i); (iii) at least 20 contiguous residues of (i); (iv)  
XX sequences that hybridize to (i), under moderately stringent conditions;  
XX (v) sequences having at least 75% or 90% identity to (i); or (vi)  
XX degenerate variants of (i). The compositions and methods of the present  
XX invention are useful for the diagnosis, prevention and/or treatment of  
XX cancer, particularly colon cancer. (I) can be used in gene therapy and  
XX (II) and (II) are useful in pharmaceutical compositions such as vaccines.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIRO  
XX at ftp.wiro.int/pub/published\_pcf\_sequences.  
XX  
XX Sequence 486 BP; 192 A; 55 C; 64 G; 175 T; 0 other;  
XX  
XX Query Match 67.0%; Score 478.2; DB 24; Length 486;  
XX Best Local Similarity 98.8%; Pred. No. 1.2e-73;  
XX Matches 480; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 CCAGAAAAAGTATTTATTTCTATTAACATCTTCTCAAGCATTTATTTATCTTA 62  
Db 1 CCAGAAAAAGTATTTATTTCTATTAACATCTTCTCAAGCATTTATTTATCTTA 60

Qy 63 TATGCTGATTTTAAAGAAATATACTTGTATTTAGAAAAACTAGAAAAAGATAAT 122  
Db 61 TATCTCACTGATTTTAAAGAAATACATTAATATTTAGAAAAAGATAAT 120  
Qy 123 GCAGATTAATTAACCTTACATGAAAAAGAAATTTAACAAGAGCTGAGAACCTTAATA 182  
Db 121 GCAGATTAATTAACCTTACATGAAAAAGAAATTTAACAAGAGCTGAGAACCTTAATA 180  
Qy 183 ATGGAATGAGATTAATTAATTTGAAACTGCACTGAAAGCAACTTTATTTGCAATTAAT 242  
Db 181 ATGGAATGAGATTAATTAATTTGAAACTGCACTGAAAGCAACTTTATTTGCAATTAAT 240  
Qy 243 NCTTAATGATGCTGTTTATGACATAATACCTGATTTTTCAGAGAAACCCATGTTAA 302  
Db 241 TCTTAATGATGCTGTTTATGACATAATACCTGATTTTTCAGAGAAACCCATGTTAA 300  
Qy 303 AATATTTTATTTTAAATAAATTAACCTGCTGTTCAAGCTGATCATATTTCTTTATTTT 362  
Db 301 AATATTTTATTTTAAATAAATTAACCTGCTGTTCAAGCTGATCATATTTCTTTATTTT 360  
Qy 363 GATTGGGAAANAATACTGTTCTGATAGCAGTGAATGCAAAATTTTATGATTTTAAAT 422  
Db 361 GATTGGGAAAGAAATACCTGTTCTGATAGCAGTGAATGCAAAATTTTATGATTTTAAAT 420  
Qy 423 CTCACATAATTTTAAATACTATTTGGAATGATTAATGACATGAGCAACACTAAT 482  
Db 421 CTCACATAATTTTAAATACTATTTGGAATGATTAATGACATGAGCAACACTAAT 480  
Qy 483 TACTGG 488  
Db 481 TACTGG 486

RESULT 5  
AAX40097  
ID AAX40097 standard; DNA; 843 BP.  
XX  
XX AAX40097;  
XX  
XX 02-JUL-1999 (first entry)  
XX  
XX Gastric cancer associated gene.  
XX  
XX Cancer associated antigen; diagnosis; research; treatment; human;  
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
XX prostate cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9904265-A2.  
XX  
XX 28-JAN-1999.  
XX  
XX 15-JUL-1998; 98WO-US14679.  
XX  
XX 22-JUN-1998; 98US-0102322.  
XX 17-JUL-1997; 97US-0896164.  
XX 10-OCT-1997; 97US-0061599.  
XX 10-OCT-1997; 97US-0061765.  
XX 10-OCT-1997; 97US-0948705.  
XX 11-OCT-1997; 97GB-0021697.  
XX  
XX (LUDW-) LUDWIG INST CANCER RES.  
XX  
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
XX Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;  
XX Tureci O;  
XX WPI; 1999-132448/11.  
XX  
XX New isolated cancer associated nucleic acids and polypeptides -  
XX isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers  
 XX  
 PS Claim 67, Page 696; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterized  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterized by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.

SO Sequence 843 BP; 237 A; 184 C; 89 G; 215 T; 118 other;

Query Match 61.1%; Score 436.2; DB 20; Length 843;  
 Best Local Similarity 73.1%; Pred. No. 2e-66;  
 Matches 516; Conservative 0; Mismatches 187; Indels 3; Gaps 3;

OY 2 GCCAGAAAAGTTATTTTATTTCTATTAACATCTTCCCAAGCATTTATCT 61  
 DB 2 GCCAAAAAATTTATTTTATTTCTATTAACATCTTCCCAAGCATTTATTT 61  
 OY 62 ATATCTCACTGAATTTTAAAGAAATTAACATTAATTAAGAAAACTAGAAAAAGATTA 121  
 DB 62 ATNNCNCNCAGNTTNNAAANNTACCTTNNNTTAAAAAACTNGAAAAAATATAT 121  
 OY 122 TGCAGATATTTAACTTACATGAAAAAATTTATTAACAAGACCTGAGAACCTTTA 181  
 DB 122 NGCAATATTTAACTTCTTCTGAAAAAATTTTTCACAAAGACGAAANCTTTT 181  
 OY 182 AATTGAATGAGATTTATTTTGAAGAACTGCACTGGAAGAACTTATTTGTTCAATTA 241  
 DB 182 ATTGAATNTAAATTTATNTTNGAAACGCGCNCNCGAAACCAANTTATGTTCAATTA 241  
 OY 242 TNCCTTAATGATGCTTTTATGACTAATACATGATTTTTCAGAGAAAGAAACCATGTTA 301  
 DB 242 TCCCTAAAGAGGNNTTTANNACCTAATNCCNGATTTTCCATATNGAAACCCNNNTTA 301  
 OY 302 AAAATATTTTATTTTAAATAAATAGCTGCTGCTCAAGCTCTGATCAATTTCTTTATTT 361  
 DB 302 AAAATNTTTTATTTTAAATAAATAGCTGCTGCTCAAGCTCTGATCAATTTCTTTATTT 361  
 OY 362 TGATTTGGGAAANAAATCTGTTCTGATGACATGAATGCAAAATTTTATGATTTTAA 421  
 DB 362 GGATTTGGGAAANAAATCTGTTCTGATGACATGAATGCAAAATTTTATGATTTTAA 421  
 OY 422 TCTCACTAATTTTAAANAACTATTGAGAAATGATTATGACATGAGCAACAACATTA 481  
 DB 422 TCTCACTAATTTTAAANAACTATTGAGAAATGATTATGACATGAGCAACAACATTA 481  
 OY 482 TTAATGCGCAGCTGTGTCATTTGTTCTTACTTATGTTCTCCCAAGGAAAACTTTAA 541  
 DB 481 TTNCGGCGCACGCGCTNGTNTTCTTACTTAACTCCCAAGGAAAACTTTAA 541  
 OY 542 ATTGAATTTGAGCAAGAAATATCTTAATATCTTTGTAAGCAAAAGAAAGCTTTT 601  
 DB 542 NCGAANTCTCCCAAAATTAACCTTAATATCTTTGTAAGCAAAAGAAAGCTTTT 601  
 OY 602 GTTTACATAGTTCTTTGGATTTTACTGTTCTTAATTTTATTTGAAACCTAATTTTACC 661  
 DB 602 GTTTACATAGTTCTTTGGATTTTACTGTTCTTAATTTTATTTGAAACCTAATTTTACC 661  
 OY 662 CCAAGACATTAATTAACATTTACTTTGTTTGCACAGTTGTTGC 707  
 DB 662 CCAAGACATTAATTAACATTTACTTTGTTTGCACAGTTGTTGC 707  
 OY 659 CCNAAACATANTTACATTTTACTGTTGTAAGGCGCAGTGTGTC 704  
 DB 659 CCNAAACATANTTACATTTTACTGTTGTAAGGCGCAGTGTGTC 704

# RESULT 6

AA119975/c  
 ID AA119975 standard; cDNA to mRNA; 351 BP.

XX AA119975;

DT 17-JUL-1996 (first entry)

XX Human gene signature HUMGS01107.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.

XX Homo sapiens.

OS W09514772-A1.

PN 01-JUN-1995.

PD 11-NOV-1994; 94WO-JP01916.

PP 12-NOV-1993; 93JP-0355504.

PR (MATS/) MATSUBARA K.  
 PA (OKUBO/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function. By preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues

XX Claim 1; Page 525; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in AA119901-126837 and which is able to hybridize to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.

XX Sequence 351 BP; 117 A; 44 C; 44 G; 141 T; 5 other;

Query Match 43.9%; Score 313.2; DB 16; Length 351;  
 Best Local Similarity 97.1%; Pred. No. 2.7e-45;  
 Matches 337; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

OY 2 GCCAGAAAAGTTATTTTATTTCTATTAACATCTTCCCAAGCATTTATCT 60  
 DB 347 GCCAGAAAAGTTATTTTATTTCTATTAACATCTTCCCAAGCATTTATTTACC 288  
 OY 61 TATATCTCACTGAATTTTAAAGAAATTAACATTAATTAAGAAAACTAGAAAAAGATTA 120  
 DB 287 TATATCTCACTGAATTTTAAAGAAATTAACATTAATTAAGAAAACTAGAAAAAGATTA 228  
 OY 121 ATGCAATTAATTAACATTTACTGAAAAAGAAATTTAAACAAGACCTGAGAACTTTAT 180  
 DB 227 ATGCAATTAATTAACATTTACTGAAAAAGAAATTTAAACAAGACCTGAGAACTTTAT 168

OY 181 AAATGAATGAGTATTAATTTGAAACATGATCGAAGAAACCACTTATTTGTCATTT 240  
 DB 167 AAATGAATGAGTATTAATTTGAAACATGATCGAAGAAACCACTTATTTGTCATTT 108  
 OY 241 ATNCTTAATGATGCTGTTTATGACTAATATACATGATTTTTCAGAGAAACCAATGTT 300  
 DB 107 ATCTTAATGATGCTGTTTATGACTAATATACATGATTTTTCAGAGAAACCAATGTT 48  
 OY 301 AAAAATATTTT-TTATTTTAAATAAGCCTGTGTTCAAGCTCTGATC 346  
 DB 47 AAAAATATTTTNNNTTTTAAATAAGCCTGTGTTCAAGCTCTGATC 1

## RESULT 7

AAH57293/C  
ID AAH57293 standard; cDNA; 262 BP.

AAH57293;

10-SEP-2001 (first entry)

Human liver specific cDNA sequence SEQ ID NO:133.

KM Human; tissue specific; diagnosis; brain; heart; skeletal muscle;  
 KM lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;  
 KM metabolic disease; developmental disease; cytostatic; immunomodulatory;  
 KM neuroprotective; gene therapy; cancer; immunopathology; neuropathology.

OS Homo sapiens.

PN WO200132927-A2.

PD 10-MAY-2001.

PF 02-NOV-2000; 2000WO-US30396.

PR 04-NOV-1999; 99US-0163508.

PA (INCY-) INCYTE GENOMICS INC.

PI Sornasse T, Seilhamer JJ, Watson GA;

DR WPI; 2001-291057/30.

PT New cell and tissue specific polynucleotides useful for diagnosis,  
 PT prognosis or monitoring of treatments for disorders where the gene is  
 PT associated with a cancer, immunopathology or neuropathology -

PS Claim 1; Page 117; 327pp; English.

CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
 CC sequences (I). (I) can have cytosstatic, immunomodulatory and  
 CC neuroprotective activities, and can be used in gene therapy. (I) and  
 CC proteins (II) encoded by then are used in high throughput screening  
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
 CC minelites, peptides, proteins, agonists, antagonists, antibodies or  
 CC their fragments, immunoglobulins, inhibitors, drug compounds and  
 CC pharmaceutical agents. Expression of (I) in a sample indicates the  
 CC differentiation of embryonic stem cells into a tissue selected from  
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
 CC tissues. (I) and (II) are used to produce an expression profile that  
 CC defines a metabolic or developmental process, treatment, condition,  
 CC disease or disorder. The gene profile can be used for diagnosis,  
 CC prognosis or monitoring of treatments and for investigating a  
 CC predisposition to a disorder where the gene is associated with a  
 CC cancer, immunopathology or neuropathology.

SO Sequence 262 BP; 103 A; 38 C; 34 G; 87 T; 0 other;

Query Match 36.2%; Score 258.4; DB 22; Length 262;  
 Best Local Similarity 98.9%; Pred. No. 7.1e-36;  
 Matches 259; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 257 TTTTATGACTATATACATGATTTTTCAGAGAAACCACTGTTAAAAATATTTTATTT 316  
 DB 252 TTTTATGACTATATACATGATTTTTCAGAGAAACCACTGTTAAAAATATTTTATTT 203  
 OY 317 TAAAAATAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTGATTTGGANAAA 376  
 DB 202 TAAAAATAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTGATTTGGANAAA 143  
 OY 377 ATACTGTTTCTGATAGCATGAAATGCAAAAATTTTATGATTTTCACTAATTTTAA 436  
 DB 142 ATACTGTTTCTGATAGCATGAAATGCAAAAATTTTATGATTTTCACTAATTTTAA 83  
 OY 437 NAACATTTGAGAAATGATTTATGACATGACATGACACATATTTACTGGCCAGCTGT 496  
 DB 82 GAACATTTGAGAAATGATTTATGACATGACATGACACATATTTACTGGCCAGCTGT 23  
 OY 497 TGGCATTTGTTTCTTACTTAG 518  
 DB 22 TGGCATTTGTTTCTTACTTAG 1

## RESULT 8

ABX83093/C  
ID ABX83093 standard; cDNA; 173 BP.

ABX83093;

24-APR-2003 (first entry)

Corn ear-derived polynucleotide (cpd) #1553.

KM Corn ear-derived polynucleotide; cpd; cDNA library; SATMON022;  
 KM SATMON023; structural gene; functional gene; regulatory gene;  
 KM corn ear-specific profile; gene transcription; gene expression;  
 KM hybrid plant; desirable trait expression; plant breeding program;  
 KM inheritance; desired characteristic; growth; development;  
 KM disease resistance; environmental adaptability; quality; yield;  
 KM multigene trait; plant; gene; ss.

OS Zea mays.

PN US6476212-B1.

PD 05-NOV-2002.

PF 14-MAY-1999; 99US-0313294.

PR 26-MAY-1998; 98US-086722P.

PA (INCY-) INCYTE GENOMICS INC.

PI Lalgudi RV, Ito LY, Sherman BK;

DR WPI; 2003-208840/20.

PT Novel purified corn-ear derived polynucleotide useful as hybridization  
 PT probe for detecting polynucleotide in sample, and for identifying,  
 PT evaluating, and altering desired characteristics associated with  
 PT growth, development -

PS Examples; SEQ ID NO 1553; 390pp; English.

CC The present invention relates to the isolation of corn ear-derived  
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries  
 CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,  
 CC functional, and regulatory genes of corn ear. The polynucleotides  
 CC sequences are useful for detecting cpds in a sample, for producing  
 CC a corn ear-specific profile of gene transcription, for detecting  
 CC altered gene expression in inbred or hybrid plants, and for screening  
 CC several molecules for specific binding to the polynucleotide. The cpds  
 CC are useful to identify, isolate, or extend identical or related  
 CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic  
 CC acid amplification or hybridisation techniques to follow the



expression of desirable traits through plant breeding programs.  
CC Preferably, the cdds are used to identify, evaluate, alter, or  
CC follow the inheritance of desired characteristics associated with  
CC growth and development, disease resistance, environmental adaptability,  
CC quality, and yield of corn. The cdds are also useful as molecular  
CC markers for studying inheritance and multigene traits in a plant  
CC breeding program. The cdds are useful for producing purified corn-ear  
CC polypeptides by recombinant techniques. They are also useful in  
CC diagnostic assays to detect or confirm conditions or diseases  
CC associated with abnormal levels of cdp expression. ABX81541-ABX89140  
CC represent corn ear-derived polynucleotides (cpds) of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site at [seqdata.uspto.gov/psipdb/entry.html](http://seqdata.uspto.gov/psipdb/entry.html).  
XX  
SQ Sequence 173 BP; 54 A; 20 C; 19 G; 80 T; 0 other;  
Query Match 23.2%; Score 165.6; DB 25; Length 173;  
Best Local Similarity 97.7%; Pred. No. 6.2e-20;  
Matches 168; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 8 AAAAGTATTATTTCTATTAACATTTCTCAAGCATTTATTTATCTATCT 67  
Db 172 AAAAGTATTATTTCTATTAACATTTCTCAAGCATTTATTTATCTATCT 113  
Qy 68 CACTGAATTTTAAGAAATTAACATTAGTATTAGAAAACTAGAAAAAATTAATGCAGA 127  
Db 112 CACGAAATTTTAAGAAATTAACATTAGTATTAGAAAACTAGAAAAAATTAATGCAGA 53  
Qy 128 TAATTAACCTTACATGAAAAAGAAATTTATTAACAAAGCATGAGACGTTA 179  
Db 52 TAATTAACCTTACATGAAAAAGAAATTTATTAACAAAGCATGAGACGTTA 1  
RESULT 9  
ID ABX40797 standard; cDNA; 397 BP.  
XX  
AC ABX40797;  
DT 20-FEB-2003 (first entry)  
XX  
DE Bovine EST associated with lactation/muscle/fat deposition #5962.  
XX  
KW Bovine; 6S; EST; expressed sequence tag; lactation; LMFD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
XX gene analysis; cattle breeding.  
XX  
OS Bos Taurus.  
XX  
FN US2002137139-A1.  
PD 26-SEP-2002.  
XX  
PF 24-SEP-2001; 2001US-0960352.  
XX  
PR 12-JAN-1999; 99US-115707P.  
PR 11-JAN-2000; 2000US-0480902.  
XX  
PA (BYAT/) BYATT J C.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
PI Byatt JC, Machialagan N, Tao N, Warren WC;  
PI WPI; 2003-110599/10.  
DR  
XX  
XX New nucleic acid associated with lactation, and muscle and fat  
XX deposition, useful for genome mapping, gene identification and  
XX analysis, cattle breeding, or for genetically improving cattle  
XX  
XX Claim 2; SEQ ID No 5962; 245bp; English.  
XX  
XX

XX  
CC The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived  
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 15112 nucleotide  
CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
CC Also included are; (1) a transformed cell having a nucleic acid  
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMFD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 15112 bovine  
CC LMFD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC [seqdata.uspto.gov/sequence.html?docid=20020137139](http://seqdata.uspto.gov/sequence.html?docid=20020137139).  
XX  
SQ Sequence 397 BP; 144 A; 75 C; 45 G; 133 T; 0 other;  
Query Match 17.7%; Score 126.2; DB 25; Length 397;  
Best Local Similarity 73.6%; Pred. No. 3.7e-13;  
Matches 204; Conservative 0; Mismatches 60; Indels 13; Gaps 3;  
Qy 435 AANAATATTGAGAAATTTGATTAATGACATGACATGACAACTAATTTACTGCCAGCT 494  
Db 14 AATTATCAGTGAAGAAATATATTAAGTCATTAATGACAACTAATTTACTGTCAGCT 73  
Qy 495 GTTGCAATGTGTTCTTACTTACTTCTCCCAAGGAAAACTTAAATTAATGAATCTTCAG 554  
Db 74 AATATGATAT-TCGTGTTCTTCTCACTCCCAAGGAAAC-----TTGAAATTCGG 124  
Qy 555 CAGAAATATCTTAAATATCTTGTAGCAAAACAAAGCTTTTGTTTACATAGTTC 614  
Db 125 CAGATATATCTTCAAAATATCATTAATTAAGCAAAATGAGAGCTTGTT-----ACATACT 180  
Qy 615 TTGGATTTTACTGTTCTTAATTTTATTCGAATCAATTTTACCCAGACCATTAAT 674  
Db 181 TTTGTATTTTTCATTTCTTAATTTTATTCGAATCAATTTTACCCAGACCATTAAT 240  
Qy 675 ACCATATTAACCTTGTGTTGACAGGTTGTTGCCAAT 711  
Db 241 ACCATATTAACCTTGTGTTGACAGGTTGTTGCCAAT 277  
RESULT 10  
ID AAQ76489 standard; DNA; 263 BP.  
XX  
AC AAQ76489;  
DT 25-MAR-2003 (updated)  
DT 23-SEP-1994 (first entry)  
XX  
DE Human genome fragment.  
XX  
XX Brain; placenta; bone marrow; genetic analysis; gene mapping;  
XX detection; homology; human; adrenal tissue; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO9401548-A2.  
XX  
XX

XX 20-JAN-1994.  
PD 13-JUL-1993; 93WO-GB01467.  
XX 13-JUL-1992; 92GB-0014857.  
XX (MED1-) MEDICAL RES COUNCIL.  
XX Gross J, Hatfield KM, Howells D, Kelly M, Shaw D;  
PI Sibson DR, Starkey M;  
XX WPI; 1994-035056/04.  
XX New nucleic acid fragment encoding gene products - can be used  
PT for genetic analysis and mapping  
XX  
XX Claim 1; Page 64; 616pp; English.  
XX Human nucleic acid fragments, isolated from brain adrenal tissue,  
CC the placenta or bone marrow comprise any of: (A) a sequence  
CC selected from (AM076401-Q77613), (B) an allelic variation of a  
CC sequence as described in (A), or (C) a sequence complementary  
CC to (A) or (B).  
CC (Updated on 25-MAR-2003 to correct PN field.)  
XX  
XX Sequence 263 BP; 97 A; 38 C; 45 G; 80 T; 3 other;  
SQ  
Query Match 15.0%; Score 107.4; DB 15; Length 263;  
Best Local Similarity 82.1%; Pred. No. 6; 3e-10;  
Matches 207; Conservative 0; Mismatches 34; Indels 11; Gaps 7;  
QY 464 TGAAGTGCACAACTAATTAATGAGCTGTCAGAGTGGATGTTCTTACTAGTCTGC 523  
DB 263 TGGAGTACACGACATGATTAATGAGCTGTCAGAGTGGATGTTCTTACTAGTCTGC 205  
QY 524 CCAAGGAAATCTTAAATGAACTTTCAGCAGAAATATCCTTAATATACTTTGTAAG 583  
DB 204 AGCA-GGAAACTCTTAATGAACTTTCAGCAGAAATATCCTTAATATACTTTGTAAG 150  
QY 584 CAAACAAAGCTTTTGTGTTACATGATGTTGGGATTTTACTGTTCTTAATTTATTT 643  
DB 149 CAAACAAAGCTTTTGTGTTACATGATGTTGGGATTTTACTGTTCTTAATTTATTT 91  
QY 644 CTGAAGTCAATTTTAC-CCGAGACCATTAATTAATTAATTTTGTGTTGACAGTTG 702  
DB 90 CTGAAGTCAATTTTACATTTTACATTTTACATTTTACATTTTACATTTTACATTTT 34  
QY 703 TTGGCAATTCGA 714  
DB 33 TATGCAATTCGA 22  
RESULT 11  
ABZ10246  
ID ABZ10246 standard; DNA; 8056 BP.  
XX  
XX ABZ10246;  
XX  
XX 16-JAN-2003 (first entry)  
XX  
XX Haematopoietic cell proliferation disorder related DNA sequence #386.  
XX  
XX Human; haematopoietic cell proliferation disorder; cytostatic;  
KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
KM cytosine methylation state; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200277272-A2.  
PN  
XX 03-OCT-2002.  
PD

PF 26-MAR-2002; 2002WO-EP03401.  
XX  
XX 26-MAR-2001; 2001US-278333P.  
PR  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Berlin K, Braun A, Distler J, Guefig D, Howe A, Mueller J;  
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Lau E;  
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;  
PI Pelet C, Schwope I, Ziebarth H;  
XX  
XX WPI; 2003-018942/01.  
XX  
XX Detecting and differentiating between hematopoietic cell proliferative  
PT disorders, comprises contacting a target nucleic acid with a reagent  
PT that distinguishes between methylated and non-methylated CpG  
PT dinucleotides -  
XX  
XX Claim 28; SEQ ID 386; 117pp; English.  
XX  
XX The present invention describes a method for detecting and  
CC differentiating between haematopoietic cell proliferative disorders  
CC associated with at least 1 gene and/or their regulatory regions in a  
CC subject. The method comprises contacting a target nucleic acid in a  
CC biological sample obtained from the subject with at least 1 reagent,  
CC which distinguishes between methylated and non-methylated CpG  
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ1118  
CC represent specifically claimed nucleotide sequences from the present  
CC invention. Oligonucleotides from the present invention can be used for  
CC differentiating between healthy haematopoietic cells and proliferative  
CC disorder haematopoietic cells; for differentiating between acute  
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
CC determining the cytosine methylation state and/or single nucleotide  
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
CC related sequences and their complements; and as primers for the  
CC amplification of haematopoietic cell proliferation disorder related  
CC DNA sequences. The nucleotide sequences from the present invention can  
CC also be used for detecting a predisposition to, or differentiation between  
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
CC haematopoietic cell proliferative disorders. The present method enables  
CC a highly specific classification of haematopoietic cell proliferative  
CC disorders allowing for improved and informed treatment of patients.  
XX  
SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;  
Query Match 10.9%; Score 77.8; DB 25; Length 8056;  
Best Local Similarity 47.3%; Pred. No. 7; 9e-05;  
Matches 338; Conservative 0; Mismatches 361; Indels 16; Gaps 3;  
QY 7 AAAAAATTTTAAATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 66  
DB 1368 AAAAAATTTTAAATTTTAAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTT 1427  
QY 67 TCACGAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
DB 1428 TTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1487  
QY 121 ATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 180  
DB 1488 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1547  
QY 181 AATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240  
DB 1548 TTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1607  
QY 241 ATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300  
DB 1608 ATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1667  
QY 301 AAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 356  
DB 1668 ATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1727

```

Oy 357 TATTTGATTTGGGAAANAAATGCTTCTGATGACGAAATGCAAAATTTTACATT 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1728 AATTTTAAATTTTAAAAATTTAAAAATTTTAAATTTAAAAATTTAAATTT 1787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 417 TTTAATCTCAATTTTAAANAATGATGAGAAATGATTAATGACATGAAGTCACAAAC 476
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1788 AATTAATGTTAATTTAATTTAATTAATTAATTAATTAATTTGTTTAAAAAAA 1847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 477 ACTAATTAATGCGCCAGCTGTTGGCATTTGTTCTTACTTACTTCTCCCAAGGAAACT 536
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1848 AAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 537 CTAAATTAATGATCTTCACAGATATATCTTAATATATCTTGTGACAAACAAAGCT 596
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1905 ---AATTTTAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTT 1961
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 597 TTTTGTGTTACATAGTCTTTGGGATTTTACTGTTCTCAATTTTATTCGAACATCAAT 656
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1962 TTTAAATTAATTAATTAATTAATTAATTTTATTTAAATTAATTAATTAATTAAT 2021
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 657 TTACCCGACCATTAATTAATTAATTAATTAATTTGTCACAGTTGTTGCCAAT 711
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2022 TAAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2076
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
ABX47851/C
ID ABX47851 standard; cDNA, 397 BP.
XX
AC ABX47851;
XX
DT 21-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #13016.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-0960352.
XX
PR 12-JAN-1999; 99US-115707P.
XX
PR 11-JAN-2000; 2000US-0480902.
XX
PA (BYATT/) BYATT J C.
PA (MATH/) MATHILAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathilagan N, Tao N, Warren WC;
XX
DR WPI, 2003-110599/10.
XX
PT New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle
XX
PS Claim 2; SEQ ID NO 13016; 245BP; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridize to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX4836-ABX49947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of

```

```

CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPRO web site:
CC seqdata.uspto.gov/sequence.html?docid=20020137139.
XX
SQ Sequence 397 BP; 124 A; 57 C; 80 G; 136 T; 0 other:
XX
Query Match 10.8%; Score 76.8; DB 25; Length 397;
Beet Local Similarity 76.4%; Pred. No. 0.00011;
Matches 120; Conservative 0; Mismatches 33; Indels 4; Gaps 2;

Oy 555 CAGAAATATCTTAAATTAATTAATTTGTAAGCAAAACAAAGCTTTTGTGTAATGATTC 614
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 397 CAGAAATGCTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 615 TTTGGATTTTACTGTTCTTAATTTTATCTGAACATCAATTTTACCCAGACCAATAT 674
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 TT--GTATTCGTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 675 ACCAATTAATCTTTGTTTNGCAGAGTTGTTGCCAAT 711
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 281 ACCAATTAATCTGTAATGACAGATGTTAGCAACT 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
ABX48095/C
ID ABX48095 standard; cDNA; 449 BP.
XX
AC ABX48095;
XX
DT 21-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #13260.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-0960352.
XX
PR 12-JAN-1999; 99US-115707P.
XX
PR 11-JAN-2000; 2000US-0480902.
XX
PA (BYATT/) BYATT J C.
PA (MATH/) MATHILAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathilagan N, Tao N, Warren WC;
XX
DR WPI, 2003-110599/10.
XX

```

PT New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 XX  
 PS Claim 2; SEQ ID NO 13260; 245bp; English.

CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMF), derived  
 CC from cattle, and the LMF nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMF nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMF nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMF EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139.

SQ Sequence 449 BP; 148 A; 70 C; 89 G; 142 T; 0 other;

Query Match 10.5%; Score 75.2; DB 25; Length 449;  
 Best Local Similarity 85.6%; Pred. No. 0.00022;  
 Matches 83; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 615 TTGGGATTTCCTGTTCTATTTTATTCGAACTCAATTTTACCCGACGCTAATT 674  
 DB 415 TTTCTATTGTTGTTTCTTAACCTTATTTCTAACTCAATTTTACCCGACGCTAATT 356  
 QY 675 ACCATATTAACTTGTGTTGACAGTGTGTTGCCAAT 711  
 DB 355 ACCATATTAACTGTGTTGACAGTGTGTTGCCAAT 319

RESULT 14

AB199762/c  
 ID AB199762 standard; cDNA; 2152 BP.

AC AB199762;

DT 07-MAR-2002 (first entry)

XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:842.

XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.

OS Mus musculus.

XX WO20018188-A2.

XX 22-NOV-2001.

XX 18-MAY-2001; 2001WO-JP04192.

XX 18-MAY-2000; 2000JP-0145977.

XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
 PI WPI; 2002-034733/04.  
 DR P-PSDB; ABB57301.  
 XX

PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or  
 PT by determining the expression profile of a gene group comprising these  
 PT genes -

PS Claim 2; Page 2087-2091; 2690bp; English.

CC The present invention describes a method for examining ischemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (1) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (1). The method  
 CC is useful for examining the ischemic condition (e.g. compressive  
 CC ischemia, occlusive ischemia or vasospastic ischemia) by measuring  
 CC expression levels of particular genes (AB199202 to AB199912, encoding  
 CC the protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The  
 CC expression levels or expression profiles produced by these genes are  
 CC used as an indicator when screening for ischemic condition-improving  
 CC drugs or therapeutics for ischemic diseases. AB199913 and AB199914  
 CC represent PCR primers for a mouse ischemic condition related sequence,  
 CC which are used in the exemplification of the present invention.

SQ Sequence 2152 BP; 586 A; 455 C; 576 G; 535 T; 0 other;

Query Match 10.3%; Score 73.4; DB 24; Length 2152;  
 Best Local Similarity 66.7%; Pred. No. 0.00045;  
 Matches 180; Conservative 0; Mismatches 82; Indels 8; Gaps 5;

QY 420 AATCTCACTAATTTTANAACTATTGAGAAATGATTATGATGACATGACACAACT 479  
 DB 2145 AATTTTATTATTAAGAACTATTAGAGAACTAGTGAATGATGAGGCACT 2086

QY 480 AATTACGCGCAGCTGTGGCATGTGTTTCTTACTTATGTTCTCCCAAGGAAACTCTT 539

DB 2085 AATTCTGCGTAACCT-ATGGAATCTGTTTCTTCTAGTCCCCCAAGGAGAA-TCCTC 2028

QY 540 AAATGATCTTCAGCAGAAATATCTTAATATTAATCTTGTAGCAAAACAAAGCTTT 599

DB 2027 ACACGAGTGTCCACACAGATGATCTCAATATATC-TTCTAGCAAAACAAAGGCTTT- 1970

QY 600 TTGTTTACATAGTCTTGTGGATTATTTACTGTCTTAAATTTTATTTCTGAACTCA-TTTT 658

DB 1969 --GTTTACTCAGTCTTCAATTTTACTAGTCTGAAATTTTATTTCTGAACTCAAGTTTA 1913

QY 659 ACCCGACACCTAATTTACATTTAATTTA 688

DB 1912 ACCCGAACCATCAATGACATGACAGCTT 1883

RESULT 15

AB210246/c  
 ID AB210246 standard; DNA; 8056 BP.

AC AB210246;

DT 16-JAN-2003 (first entry)

XX Haematopoietic cell proliferation disorder related DNA sequence #386.

XX Human; haematopoietic cell proliferation disorder; cytostatic;  
 XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 XX cytosine methylation state; gene; ds.

XX Homo sapiens.

XX WO20027272-A2.

XX

XX

PD 03-OCT-2002.  
 XX 26-MAR-2002; 2002MO-EP03401.  
 PF 26-MAR-2001; 2001US-278333P.  
 PR (EPIC-) EPIGENOMICS AG.  
 PA Berlin K, Braun A, Dietler J, Guetig D, Howe A, Mueller J;  
 PI Olex A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu B;  
 P1 Levin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;  
 P1 Pelet C, Schoepe I, Ziebarth H;  
 XX WPI, 2003-018942/01.  
 DR  
 XX  
 PT Detecting and differentiating between hematopoietic cell proliferative  
 PT disorders, comprises contacting a target nucleic acid with a reagent  
 PT that distinguishes between methylated and non-methylated CpG  
 PT dinucleotides -  
 XX  
 PS Claim 28; SEQ ID 386; 117pp; English.  
 XX  
 CC The present invention describes a method for detecting and  
 CC differentiating between haematopoietic cell proliferative disorders  
 CC associated with at least 1 gene and/or their regulatory regions in a  
 CC subject. The method comprises contacting a target nucleic acid in a  
 CC biological sample obtained from the subject with at least 1 reagent,  
 CC which distinguishes between methylated and non-methylated CpG  
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118  
 CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used: for  
 CC differentiating between healthy haematopoietic cells and proliferative  
 CC disorder haematopoietic cells; for differentiating between acute  
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of haematopoietic cell proliferation disorder related  
 CC DNA sequences. The nucleotide sequences from the present invention can  
 CC also be used for detecting a predisposition to, differentiation between  
 CC subtypes, diagnosis, prognosis, treatment and/or monitoring of  
 CC haematopoietic cell proliferative disorders. The present method enables  
 CC a highly specific classification of haematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients.  
 CC  
 XX  
 SO Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;  
 Query Match 9.9%; Score 71; DB 25; Length 8056;  
 Best Local Similarity 47.0%; Pred. No. 0.0012;  
 Matches 283; Conservative 0; Mismatches 313; Indels 6; Gaps 2;  
 QY 18 TTAATTTCTAATAAATCTTCTCAAGCAATTTATCTATATCTCAGTAATTT 77  
 DB 1650 TTTTATTTTAAATTTTCAATTTTCAATTTTAAATTTTAAATTAATATAT 1591  
 QY 78 TAAGAAATPACCTTGTATAGAAAAGTAAAGTAAATGATGATTAATTAAC 137  
 DB 1590 AAATTAATTAATTAATTAATTAATTTTAAATTAATTTTAAATTAATTTT 1531  
 QY 138 TACATGAAGAAAGAAATTAATCAAGAGTGAAGCTTAATTAATGAATGAGATTA 197  
 DB 1530 TTTTAAATTAATTTATTTTATATATTAATTAATTAATTAATTAATTAATTA 1471  
 QY 198 TAAATTTGAAGCTGATCGAAGCAATTTATTTGTCATTAATTAATGATGATG 257  
 DB 1470 TAAATTTTAAATTAATTAATTTTATTTTAAATTAATTAATTAATTAATTA 1411  
 QY 258 TTTATGACTAATACCTGATTTTCAAGAGAAAGCCATGTTAAATTAATTTTATTTT 317  
 DB 1410 AAAAAAATCAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTA 1351  
 QY 318 AAAAAATAGCCGTGTGTCAGCTGATCATATTTCTTTATTTT---GATTTGGGAAN 373

DB 1350 TAAATTAATTTTAAATTAATTAATTAATTTTATTAATTAATTTTAAATTA 1291  
 QY 374 AAATACGTTTCTGATGACATGAATGCAAAATTTTATGATTTTATCTCAGTAATTT 433  
 DB 1290 ATATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1231  
 QY 434 TAAATTAATTTGAGAAATTTGA--TTAATGACATGAGTGCACAACTAATTAATGAGCA 491  
 DB 1230 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1171  
 QY 492 GCTGTGGCATTTGTTCTTACTTACTTACTTCCCAAGGAAAGCTTAATTAATGATTT 551  
 DB 1170 TCAAAATATATTAATTTTAAATTTTCAAAATATCAAAATATTAATTAATTAATTA 1111  
 QY 552 CAGCAGATTAATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 611  
 DB 1110 TAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1051  
 QY 612 TT 613  
 DB 1050 TT 1049

Search completed: November 27, 2003, 06:23:42  
 Job time : 236.619 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 06:06:02 ; Search time 269.238 Seconds  
(without alignments)  
8737.270 Million cell updates/sec

Title: US-09-835-992A-19

Perfect score: 714

Sequence: 1 cgcagcaaaaagtcattcttca.....cacagtcgtcccaattca 714

Scoring table:

Gapop 10.0, Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710	99.4	714	9	US-09-835-992A-19 Sequence 19, Appl
2	672.8	94.2	1052	10	US-09-880-107-3296 Sequence 3296, Ap
3	611.8	85.7	687	9	US-09-835-992A-20 Sequence 20, Appl
4	483.4	67.6	506	10	US-09-796-692-7228 Sequence 7228, Ap
5	482.4	67.6	506	14	US-10-040-862-7228 Sequence 7228, Ap
6	478.2	67.0	486	10	US-09-998-558-578 Sequence 578, App
7	436.2	61.1	843	9	US-09-835-992A-22 Sequence 22, Appl
8	188	26.3	510	14	US-10-102-524-161 Sequence 161, App
9	188	26.3	510	14	US-10-102-524-188 Sequence 188, App
10	188	26.3	510	14	US-10-102-524-645 Sequence 645, App
11	125.2	17.7	397	10	US-09-960-352-5962 Sequence 5962, Ap
12	76.8	10.8	397	10	US-09-960-352-13016 Sequence 13016, A
13	75.2	10.5	449	10	US-09-960-352-13260 Sequence 13260, A
14	67.6	9.5	960	14	US-10-198-846-6381 Sequence 6381, Ap
15	67.4	9.4	4985	12	US-10-056-405-10 Sequence 10, Appl
16	67.4	9.4	4985	14	US-10-094-240-10 Sequence 10, Appl

17	65.2	9.1	7676	12	US-10-240-485-151	Sequence 151, App
18	65.2	9.1	3673778	12	US-10-312-841-1	Sequence 1, Appl
19	64.6	9.0	6105	12	US-10-311-455-1445	Sequence 1445, Ap
20	63.6	8.9	6944	14	US-10-172-086-111	Sequence 111, App
21	62.2	8.7	291	10	US-09-960-352-13694	Sequence 13694, A
22	62	8.7	3673778	12	US-10-312-841-2	Sequence 2, Appl
23	61.8	8.7	5962	12	US-10-311-455-1260	Sequence 1260, Ap
24	61.4	8.6	17721	12	US-10-311-455-1702	Sequence 1702, Ap
25	61.4	8.6	3673778	12	US-10-312-841-1	Sequence 1, Appl
26	60.8	8.5	424	10	US-09-960-352-11218	Sequence 11218, A
27	60.6	8.5	10467	12	US-10-240-453-327	Sequence 327, App
28	59.6	8.3	6912	12	US-10-240-453-245	Sequence 245, App
29	59.6	8.3	7631	12	US-10-311-455-834	Sequence 834, App
30	59.4	8.3	12138	12	US-10-311-455-1916	Sequence 1916, Ap
31	59.4	8.3	12138	12	US-10-240-453-210	Sequence 210, App
32	59.2	8.3	7046	12	US-10-311-455-2090	Sequence 2090, App
33	58.8	8.2	113515	12	US-10-311-455-2147	Sequence 2147, Ap
34	58.6	8.2	480	10	US-09-960-352-5301	Sequence 5301, Ap
35	58.4	8.2	10151	12	US-10-311-455-2405	Sequence 2405, Ap
36	58.4	8.2	16724	12	US-10-311-455-1064	Sequence 1064, Ap
37	58.4	8.2	16724	12	US-10-240-485-90	Sequence 90, Appl
38	58.4	8.2	11515	12	US-10-311-455-2148	Sequence 2148, Ap
39	58.2	8.2	640681	10	US-09-790-988-1	Sequence 1, Appl
40	58	8.1	3673778	12	US-10-312-841-2	Sequence 2, Appl
41	57.8	8.1	539	14	US-10-198-846-1369	Sequence 1369, Ap
42	57.8	8.1	6389	12	US-10-311-455-1933	Sequence 1933, Ap
43	57.6	8.1	3991	14	US-10-074-045-60	Sequence 60, Appl
44	57.6	8.1	6361	12	US-10-311-455-1114	Sequence 1114, Ap
45	57.6	8.1	6609	12	US-10-311-455-1855	Sequence 1855, Ap

## ALIGNMENTS

RESULT 1  
US-09-835-992A-19  
Sequence 19, Application US/09835992A  
Patent No. US20020037541A1  
GENERAL INFORMATION:  
APPLICANT: Obata, Yuichi  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER  
FILE REFERENCE: L0461/7112  
CURRENT APPLICATION NUMBER: US/09/835, 992A  
CURRENT FILING DATE: 2001-04-16  
PRIOR APPLICATION NUMBER: US 08/896, 164  
PRIOR FILING DATE: 1997-07-17  
NUMBER OF SEQ ID NOS: 87  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 19  
LENGTH: 714  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Unsure  
LOCATION: (243)..(243)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (373)..(373)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (437)..(437)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (702)..(702)  
OTHER INFORMATION: n = a, c, g or t  
US-09-835-992A-19

Query Match 99.4%; Score 710; DB 9; Length 714;  
Best Local Similarity 100.0%; Pred. No. 7.5e-118;  
Matches 714; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGCAGCAAAAAGTATTATTATTTCTATTAAACATCTCTCAAGCATATTATTATCC 60

```
Db 1 CGCCGAAAAGTATTTTAAATTTCTATTAACATTCCTCCAAAGCATATTTATCC 60
Qy 61 TATATCTCATGTAATTTTAAAGAAATPACCTTGTATTTGAAAAAATAGAAAAAGATPA 120
Db 61 TATATCTCATGTAATTTTAAAGAAATPACCTTGTATTTGAAAAAATAGAAAAAGATPA 120
Qy 121 ATGCGATPATAAATCTTACATGAAAAAGAAAAATTTATPACAAAGACCTGAAAGCTTAT 180
Db 121 ATGCGATPATAAATCTTACATGAAAAAGAAAAATTTATPACAAAGACCTGAAAGCTTAT 180
Qy 181 AAATGAAATGAGATTAATTAATTTGAAAACTGCATCTGAAAGCAACTTATTTGTCATTT 240
Db 181 AAATGAAATGAGATTAATTAATTTGAAAACTGCATCTGAAAGCAACTTATTTGTCATTT 240
Qy 241 ATNCTTAATGATGTTGTTTATGACTAATPACATGATTTTTCAGAAAGAAACCATGTT 300
Db 241 ATNCTTAATGATGTTGTTTATGACTAATPACATGATTTTTCAGAAAGAAACCATGTT 300
Qy 301 AAAATATTTTATTTTAAATAAGCCCTGCTGATGCTGATCATATTTCTTTTATTT 360
Db 301 AAAATATTTTATTTTAAATAAGCCCTGCTGATGCTGATCATATTTCTTTTATTT 360
Qy 361 TTGATTTGGGAAANAATACTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTAA 420
Db 361 TTGATTTGGGAAANAATACTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTAA 420
Qy 421 ATCTCATPATAATTTTAAATCTATTTGAAATGATTAATGACATGAGTGCACAACTTA 480
Db 421 ATCTCATPATAATTTTAAATCTATTTGAAATGATTAATGACATGAGTGCACAACTTA 480
Qy 481 ATTACTGGCAGCTGTGGCATTTGTTCTTACTTACTAGTCTCCCAAGGAAAACTCTTA 540
Db 481 ATTACTGGCAGCTGTGGCATTTGTTCTTACTTACTAGTCTCCCAAGGAAAACTCTTA 540
Qy 541 AATTGAATCTTCAGAGAAATATCCTTAATATPACTTTGTGAAGCAAAACAAAGCTTTT 600
Db 541 AATTGAATCTTCAGAGAAATATCCTTAATATPACTTTGTGAAGCAAAACAAAGCTTTT 600
Qy 601 TGTTCACATAGTCTTGGGATTTTACTGTTCCATTTTATTTTCTGAAACTCAATTTTAC 660
Db 601 TGTTCACATAGTCTTGGGATTTTACTGTTCCATTTTATTTTCTGAAACTCAATTTTAC 660
Qy 661 CCCAGACATATTAATCAATTAATCTTGTGTTGACAGTGTGGCAATTC 714
Db 661 CCCAGACATATTAATCAATTAATCTTGTGTTGACAGTGTGGCAATTC 714

RESULT 2
US-09-880-107-3296/C
; Sequence 3296, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darc1 T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3296
; LENGTH: 1052
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U11313
```

```
US-09-880-107-3296
Query Match 94.2%; Score 672.8; DB 10; Length 1052;
Best Local Similarity 98.2%; Pred. No. 4e-111;
Matches 699; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

Qy 2 GCCAGAAAAGTATTTTAAATTTCTATTAACATTCCTTCAAGCATATTTATTCCT 61
Db 1027 GCCAGAAAAGTATTTTAAATTTCTATTAACATTCCTTCAAGCATATTTATTCCT 968
Qy 62 ATATCTCATGTAATTTTAAAGAAATPACCTTGTATTTGAAAAAATAGAAAAAGATPA 121
Db 967 ATATCTCATGTAATTTTAAAGAAATPACCTTGTATTTGAAAAAATAGAAAAAGATPA 908
Qy 122 TGCAGATATAAATCTTACATGAAAAAGAAAAATTTATPACAAAGACCTGAAAGCTTATA 181
Db 907 TGCAGATATAAATCTTACATGAAAAAGAAAAATTTATPACAAAGACCTGAAAGCTTATA 848
Qy 182 AATTGAAATGAGATTAATTAATTTGAAAACTGCATCTGAAAGCAACTTATTTGTCATTT 241
Db 847 AATTGAAATGAGATTAATTAATTTGAAAACTGCATCTGAAAGCAACTTATTTGTCATTT 788
Qy 242 TNCCTTAATGATGTTGTTTATGACTAATPACATGATTTTTCAGAAAGAAACCATGTTA 301
Db 787 TNCCTTAATGATGTTGTTTATGACTAATPACATGATTTTTCAGAAAGAAACCATGTTA 728
Qy 302 AAAATATTTTATTTTAAATAAGCCCTGCTGATGCTGATCATATTTCTTTTATTT 361
Db 727 AAAATATTTTATTTTAAATAAGCCCTGCTGATGCTGATCATATTTCTTTTATTT 668
Qy 362 TGATTTGGGAAANAATACTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTAA 421
Db 667 TGATTTGGGAAANAATACTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTAA 608
Qy 422 TCTCATPATAATTTTAAATCTATTTGAAATGATTAATGACATGAGTGCACAACTTA 481
Db 607 TCTCATPATAATTTTAAATCTATTTGAAATGATTAATGACATGAGTGCACAACTTA 548
Qy 482 TTACTGGCAGCTGTGGCATTTGTTCTTACTTACTAGTCTCCCAAGGAAAACTCTTA 541
Db 547 TTACTGGCAGCTGTGGCATTTGTTCTTACTTACTAGTCTCCCAAGGAAAACTCTTA 489
Qy 542 AATTGAATCTTCAGAGAAATATCCTTAATATPACTTTGTGAAGCAAAACAAAGCTTTT 601
Db 488 AATTGAATCTTCAGAGAAATATCCTTAATATPACTTTGTGAAGCAAAACAAAGCTTTT 429
Qy 602 GTTTACATAGTCTTGGGATTTTACTGTTCCATTTTATTTCTGAAACTCAATTTTAC 661
Db 428 GTTTACATAGTCTTGGGATTTTACTGTTCCATTTTATTTCTGAAACTCAATTTTAC 370
Qy 662 CCAAGACATATTAATCAATTAATCTTGTGTTGACAGTGTGGCAATTC 713
Db 369 CCAAGACATATTAATCAATTAATCTTGTGTTGACAGTGTGGCAATTC 318

RESULT 3
US-09-835-992a-20
; Sequence 20, Application US/09835992A
; Patent No. US20020037541A1
; GENERAL INFORMATION:
; APPLICANT: Obata, Yuichi
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER
; FILE REFERENCE: L0461/7112
; CURRENT APPLICATION NUMBER: US/09/835,992A
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: US 08/896,164
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 687
; TYPE: DNA
```



ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Unsure  
LOCATION: (67)..(67)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (105)..(105)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (412)..(412)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (541)..(541)  
OTHER INFORMATION: n = a, c, g or t  
US-09-835-992a-20

Query Match 85.7%; Score 611.8; DB 9; Length 687;  
Best Local Similarity 97.1%; Pred. No. 2.9e-100;  
Matches 671; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

15 ATTTAATTTCTATTAAACATCTCTCAAGCTTATTTATCTCTCAAGCTTAA 74  
1 ATTTAATTTCTATTAAACATCTCTCAAGCTTATTTATCTCTCAAGCTTAA 60  
75 TTTAAGAAATTAACATCTAGTATTAGAAAACAGAAAAGATAATGACATATTA 134  
61 TTTAAGAAATTAACATCTAGTATTAGAAAACAGAAAAGATAATGACATATTA 120  
135 ACTTACATGAAAGAAAGAAATTTATTAACAAGACTGAGACGTTAATTAATGAAATGAGA 194  
121 ACTTACATGAAAGAAAGAAATTTATTAACAAGACTGAGACGTTAATTAATGAAATGAGA 180  
195 TTATTAATTTGAAAACATGATCTGAAAAGCAATTTATTTGCAATTAATTAATGATG 254  
181 TTATTAATTTGAAAACATGATCTGAAAAGCAATTTATTTGCAATTAATTAATGATG 240  
255 TGTATTATGACATTAATACATGATTTTTCAGAGGAAACCATGTTAAATATTTAT 314  
241 TGTATTATGACATTAATACATGATTTTTCAGAGGAAACCATGTTAAATATTTAT 300  
315 TTTAAGAAATTAACATCTAGTATTAGAAAACAGAAAAGATAATGACATATTA 374  
301 TTTAAGAAATTAACATCTAGTATTAGAAAACAGAAAAGATAATGACATATTA 360  
375 AAATACATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 434  
361 AAATACATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 420  
435 AAATACATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 494  
421 AAATACATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 480  
495 GTTGCATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 554  
481 GTTGCATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 539  
555 CAGAAATTAATCTTAAATTAATCTTTGTAAGCAAAAGCTTTTGTGTTACATAGTTC 614  
540 CAGAAATTAATCTTAAATTAATCTTTGTAAGCAAAAGCTTTTGTGTTACATAGTTC 597  
615 TTTGGATTTATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 674  
598 TTTGGATTTATGATTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTCAATTTT 656  
675 ACCATA-TTAACTTTGTTTGTGCAAGTGT 704  
657 ACCATA-TTAACTTTGTTTGTGCAAGTGT 687

RESULT 4  
US-09-796-692-7228  
Sequence 7228; Application US/09796692  
Publication No. US20020198362A1

GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Aigster, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077, 001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 7228  
LENGTH: 506  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-7228

Query Match 67.6%; Score 482.4; DB 10; Length 506;  
Best Local Similarity 99.2%; Pred. No. 3.7e-77;  
Matches 483; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

2 GCCAGAAAAGTATTTTATTTCTATTAACAATCTTCTCAAGCAATTAATTTATCT 61  
20 GCCAGAAAAGTATTTTATTTCTATTAACAATCTTCTCAAGCAATTAATTTATCT 79  
62 ATATCTCACTGAATTTTAAAGAAATTAACATTTAGTATTAAGAAAACCTGAGAAATATA 121  
80 ATATCTCACTGAATTTTAAAGAAATTAACATTTAGTATTAAGAAAACCTGAGAAATATA 139  
122 TGCAGATTAATTAACATTAACATGAAGAAATTAATTAACAAGAACTGAGAAATATA 181  
140 TGCAGATTAATTAACATTAACATGAAGAAATTAATTAACAAGAACTGAGAAATATA 199  
182 AATGAAGATTAATTAATTTGAAAACCTGATCTGAAAAGCAATTTATTTGTTCAATTA 241  
200 AATGAAGATTAATTAATTTGAAAACCTGATCTGAAAAGCAATTTATTTGTTCAATTA 259  
242 TNCCTTAATGATGATTTTATGACTAATACATGATTTTCAAGAAAGAAACCTGATTA 301  
260 TTTTAAATGATGATTTTATGACTAATACATGATTTTCAAGAAAGAAACCTGATTA 319  
302 AAATATTTTATTTTAAAGAAAGCTGATCTGATCAATTTCTTTATTTT 361  
320 AAATATTTTATTTTAAAGAAAGCTGATCTGATCAATTTCTTTATTTT 379  
362 TGAATTTGGAAGAAATTAATCTGTTCTGATGACATGAAGAAATTTTATTTTAA 421  
380 TGAATTTGGAAGAAATTAATCTGTTCTGATGACATGAAGAAATTTTATTTTAA 439



Db 301 AAATTTTATTTTAAATTAAGCTGTTCAAGCTCTGATCATTTCTTTATTTT 360  
Qy 363 GATTGGGANAATAATCTGTTTCTGATGACATGAATGCAAAATTTTATTTTAT 422  
Db 361 GATTGGGANAATAATCTGTTTCTGATGACATGAATGCAAAATTTTATTTTAT 420  
Qy 423 CTCCTAATTTTAAACATTTGAGAAATGATTAATGACATGAGTGCAACACTAAT 482  
Db 421 CTCCTAATTTTAAACATTTGAGAAATGATTAATGACATGAGTGCAACACTAAT 480  
Qy 483 TACTGG 488  
Db 481 TACTGG 486

RESULT 7  
US-09-835-992A-22  
Sequence 22, Application US/09835992A  
Patent No. US20020037541A1  
GENERAL INFORMATION:  
APPLICANT: Obata, Yuichi  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND  
METHODS FOR DIAGNOSING AND TREATING GASTRIC CANCER  
FILE REFERENCE: L0461/7112  
CURRENT APPLICATION NUMBER: US/09/835,992A  
CURRENT FILING DATE: 2001-04-16  
PRIOR APPLICATION NUMBER: US 08/896,164  
PRIOR FILING DATE: 1997-07-17  
NUMBER OF SEQ ID NOS: 87  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 22  
LENGTH: 843  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Unsure  
LOCATION: (12)..(12)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (19)..(19)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (33)..(33)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (35)..(35)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (42)..(42)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (47)..(47)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (56)..(56)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (64)..(65)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (69)..(69)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (71)..(71)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (74)..(74)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (78)..(78)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure

LOCATION: (80)..(80)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (84)..(84)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (92)..(93)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (95)..(95)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (107)..(107)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (122)..(122)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (130)..(130)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (139)..(139)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (148)..(148)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (157)..(157)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (164)..(164)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (169)..(169)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (174)..(174)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (176)..(176)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (179)..(179)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (185)..(185)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (189)..(189)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (191)..(191)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (200)..(200)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (203)..(203)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (208)..(208)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (213)..(214)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (216)..(216)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (225)..(225)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (229)..(229)

1	OTHER INFORMATION: n = a, c, g or t
2	NAME/KEY: Unsure
3	LOCATION: (246)..(246)
4	OTHER INFORMATION: n = a, c, g or t
5	NAME/KEY: Unsure
6	LOCATION: (249)..(249)
7	OTHER INFORMATION: n = a, c, g or t
8	NAME/KEY: Unsure
9	LOCATION: (255)..(256)
10	OTHER INFORMATION: n = a, c, g or t
11	NAME/KEY: Unsure
12	LOCATION: (260)..(260)
13	OTHER INFORMATION: n = a, c, g or t
14	NAME/KEY: Unsure
15	LOCATION: (262)..(263)
16	OTHER INFORMATION: n = a, c, g or t
17	NAME/KEY: Unsure
18	LOCATION: (270)..(270)
19	OTHER INFORMATION: n = a, c, g or t
20	NAME/KEY: Unsure
21	LOCATION: (274)..(274)
22	OTHER INFORMATION: n = a, c, g or t
23	NAME/KEY: Unsure
24	LOCATION: (287)..(287)
25	OTHER INFORMATION: n = a, c, g or t
26	NAME/KEY: Unsure
27	LOCATION: (292)..(292)
28	OTHER INFORMATION: n = a, c, g or t
29	NAME/KEY: Unsure
30	LOCATION: (296)..(296)
31	OTHER INFORMATION: n = a, c, g or t
32	NAME/KEY: Unsure
33	LOCATION: (305)..(305)
34	OTHER INFORMATION: n = a, c, g or t
35	NAME/KEY: Unsure
36	LOCATION: (307)..(307)
37	OTHER INFORMATION: n = a, c, g or t
38	NAME/KEY: Unsure
39	LOCATION: (312)..(312)
40	OTHER INFORMATION: n = a, c, g or t
41	NAME/KEY: Unsure
42	LOCATION: (329)..(329)
43	OTHER INFORMATION: n = a, c, g or t
44	NAME/KEY: Unsure
45	LOCATION: (332)..(332)
46	OTHER INFORMATION: n = a, c, g or t
47	NAME/KEY: Unsure
48	LOCATION: (342)..(342)
49	OTHER INFORMATION: n = a, c, g or t
50	NAME/KEY: Unsure
51	LOCATION: (348)..(348)
52	OTHER INFORMATION: n = a, c, g or t
53	NAME/KEY: Unsure
54	LOCATION: (357)..(357)
55	OTHER INFORMATION: n = a, c, g or t
56	NAME/KEY: Unsure
57	LOCATION: (379)..(379)
58	OTHER INFORMATION: n = a, c, g or t
59	NAME/KEY: Unsure
60	LOCATION: (381)..(381)
61	OTHER INFORMATION: n = a, c, g or t
62	NAME/KEY: Unsure
63	LOCATION: (387)..(388)
64	OTHER INFORMATION: n = a, c, g or t
65	NAME/KEY: Unsure
66	LOCATION: (394)..(395)
67	OTHER INFORMATION: n = a, c, g or t
68	NAME/KEY: Unsure
69	LOCATION: (399)..(400)
70	OTHER INFORMATION: n = a, c, g or t
71	NAME/KEY: Unsure
72	LOCATION: (406)..(406)

Query Match	61.1%	Score 436.2	DB 9	Length 843
Best Local Similarity	73.1%	Pred. No. 8e-69		
Matches 516	Conservative 0	Matches 187	Indels 3	Gaps 3
QY	2	GGCAGAAAAAGTATTTTAAATTTCTATTAACATTTCTTCAAGACATTAATTTATCT	61	
DB	2	GGCAAAAAAANTTATTTAAATTTTCTATTAANCNTCTCCCAANCAATTTATTNACCT	61	
QY	62	ATATCTCACTGAAATTTTAAAGAAATTAACATTAGTATTGAAAAACCTAGAAAAAAGATATA	121	
DB	62	ATTNNCNCNGANTTTTNAANAANTACCTTTNNTTTAAAAAACCTGGAAAAAATAATAT	121	
QY	122	TGCAGATTAATTAACCTCATGAAAAAGAAAAATTTATACAAAGAGACTGAGACGTTATA	181	
DB	122	NGCAAAATANTTAACTTTCCTGAAAAANGAAATTTNTTACCAANGAGACGAAANCNTTATA	181	
QY	182	AATTGAATGAGATTATATAATTGAAAAATGATCTGAAGAAACCTTATTTGTCATTA	241	
DB	182	ATTNGAANTTAAATTTATNTTNGAANNGCNCNGAAACCAACTTATATGCTCAATTA	241	
QY	242	TNCTTAATGATGGTGTATTATGACTTAATACACTGATTTTCAAGAGAAACCATGTTA	301	
DB	242	TCCTTAANAGAGGANNTTTANANACTAATNCCNGATTTTCCAAATANGAANCCNNNTTA	301	
QY	302	AAATATTTTATTTTAAAAATTAAGCCTGTGTCAAGCTGTATCATATTTCTTTATTT	361	
DB	302	AAANTNTTTTAAATTTTAAAAATTAACCCGNTTCCAAACCCCGATCANATTCCTTNAATTT	361	
QY	362	TGATTTGGGAANAATAATACGTCTCTCATATGACATGAATAATGAAAAATTTTATGATTTAA	421	
DB	362	GGATTTGGGAAAAAATATNCGTTTCNNNTACCCNNAANGCAANNTTTTAAATTTTAA	421	
QY	422	TCTCACTAATTTTAAANACTATTAGAAATTTGATTAAATGACATGAGTGACAACTAA	481	
DB	422	CCCCCTANTTTTAAANCTATENGAANAANTNGATTANNGACTGAATTCG-CAACCTAN	480	
QY	482	TTACTGCGCAGCTGTGGCATTGTGTTCTTACTTACTTCTCCAGAGAAAACTTTAA	541	
DB	481	TTNCGGCGCACNCGTGCGCNGTNTTCTTACTTANTCCGCCCA-AGAAANNCTTTAA	539	
QY	542	ATTGAATCTTCAGAGAAATATCCTTAAATATCTTTGTAAGCAAAACAAAGCTTTT	601	
DB	540	NCNGAANTCNCCAAATTAACCTTAAATATCCTTGTATACCAAAACAAACCTTTT	599	
QY	602	GTTTACATAGTCTTTGGGATTTTACTGTTCTTATTTATTTGAAACTCAATTTTACC	661	
DB	600	GTTTACNTA-NTCCTTGGGATTTAACGGGTCCCAATTTNATCNGAACCACTTTTCC	658	
QY	662	CCAGACCATTAATTAACATTAATTACTTGTNTTGACAGTTGTTTC	707	
DB	659	CNNAACCATANTTACATTTTACTTGTGTAAGCNCGTGTGTTGC	704	
RESULT 8				
US-10-102-524-161/c				
; Sequence 161, Application US/10102524				
; Publication No. US20030109434A1				
; GENERAL INFORMATION:				
; APPLICANT: Algate, Paul A.				
; APPLICANT: Mannion, Jane				
; APPLICANT: Gaiger, Alexander				
; APPLICANT: Gordon, Brian				
; APPLICANT: Harlocker, Susan L.				
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER				
; FILE REFERENCE: 210121.572				
; CURRENT APPLICATION NUMBER: US/10/102,524				
; CURRENT FILING DATE: 2002-03-19				
; NUMBER OF SEQ ID NOS: 1863				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 161				
; LENGTH: 510				
; TYPE: DNA				



```

? SEQ ID NO: 13260
?
? LENGTH: 449
?
? TYPE: DNA
?
? ORGANISM: Bos taurus
?
? OTHER INFORMATION: Clone ID: 57-BOVMS1-016-Q1-E1-G1
US-09-960-352-13260

```

187 AATGAGATTATATTTGAAACTGCATCTGAAGCAAACTTATTGTTCAATTATNCTT 246

Db	810	AAANNTTNTAANTTATTAAACCAAAATTTTTTTTTTAAAAAAATTTTTTAAANNTAAAT	751
Qy	247	AATGATGCGTATTATGACTAATPACGTCGATTTTTCAGAGAGAAACCATGTTAAAAAT	306
Db	750	AAATTNAATTTTTTAAATNAAAAAATTTNAAATTTTAA-----NCANNT	699
Qy	307	ATTTTATTATTTAAAAATAAGCCGTCGTCAGCTGCATATTTCTTTATTTTGATT	366
Db	698	TTTTTTTTNTNTNNATPAAAAAANNTTTTAAATPAAAAAANNNATTTTTTTTANNAATA	639
Qy	367	TGGGAANAAAAATACGTGTTCTGATPAGACATGAATGCMAATTTTATGATTTTAAATCTCA	426
Db	638	TTTAAAAAANNNATTTTTTTTTTTTATANTTATATAAAAATTTTATTTNTNTTTCTNA	579
Qy	427	CTAATTTTAAUNACTATGAGAAATGATTAATGACATGAGTCAACAACACTAAATTAAT	486
Db	578	AAANAAAAAATTTTTTTTNAANNTTTTTTAAACCTTNAATTNAAANNAATTTTTTTT	519
Qy	487	GGCCAGCTGTGGCAATTGTCGTTCTTACTAGTCTCCAGGAAAACTCTTAAATTGA	546
Db	518	NNNTTNNANATPAAAAAATTTATTTTTTTTANTTCTATAAATTPAAAAANNAANNAATATA	459
Qy	547	ATCTTCAGACGATATATCCTTAATATATACTTTGTGAGAAAAACAAA	593
Db	458	ATAATTTTNAAAAAATAATTTTATANTTATVNNNAANNAATTTTAAAA	412

```

RESULT 15
US-10-056-405-10
; Sequence ID: Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFATORY GENES, POLYPEPTIDES, AND METHODS OF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
; US-10-056-405-10

```

Query Match	9.4%	Score 67.4;	DB 12;	Length 4985;
Best Local Similarity	44.4%	Pred. No. 0.014;		
Matches 260; Conservative	0;	Mismatches 325;	Indels 0;	Gaps 0;

QY	128	TAATTAACCTACATGAAAAAGGAAAAATTATPAACAAAGACCTGGAACGTTATAAATTGA	187
QY	128	TAATTAACCTACATGAAAAAGGAAAAATTATPAACAAAGACCTGGAACGTTATAAATTGA	187
Db	2277	TGATTTTATTTTAAAGAGAAAGATPAACATTTTATGTAACCAATTTTGTTACAGAAC	2333
QY	188	AATGAGATTATTAATTGAAAACTGCATCTGAAAGCAACCTTATTTGTTCAATTATCTTA	247
Db	2337	AAATATACGAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA	2398
QY	248	ATGATGCGCTTTTATGACTATATPACACTGATTTTTCAGAGAGAAACCATGTTAAATA	307
Db	2397	TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA	2456
QY	308	TTTTTATTTTAAAAATAAGCCTGCTGTCAGCCTGATCATATTTCTTTTAATTTGATTT	367
Db	2457	TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA	2516
QY	368	GGGAAATAAATACGCTTCTGATGACATGAATGCAAAATTTTGAATTTTAAATCTCAC	427
Db	2517	TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA	2578
QY	428	TAATTTTAAATACTATTGAGAAATTGATTAATGACATGAAGTGCACCAACTAATTTACTG	487

[illegible]

Search completed: November 27, 2003, 12:52:53  
Job time : 275.238 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using SW model

Run on: November 27, 2003, 06:00:22 ; Search time 52.701 Seconds  
(without alignment)  
5979.910 Million cell updates/sec

Title: US-09-835-992A-19  
Perfect score: 714  
Sequence: 1 CGCGAGAAAAGTATCTTCA.....CACAGTCTGTCGCAATCA 714

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCUTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	710	99.4	714	US-08-896-164-19	Sequence 19, Appl
2	611.8	85.7	687	US-08-896-164-20	Sequence 20, Appl
3	436.2	61.1	843	US-08-896-164-22	Sequence 22, Appl
4	165.6	23.2	173	US-09-313-299A-1553	Sequence 1553, Ap
5	66.4	9.3	5852	US-07-867-106-2	Sequence 2, Appl
6	61.2	8.6	20674	US-09-641-638-651	Sequence 651, App
7	57.4	8.0	927	US-09-134-001C-150	Sequence 150, App
8	55.2	7.7	837	US-08-998-416-288	Sequence 288, App
9	53	7.4	19124	US-08-487-826B-13	Sequence 13, Appl
10	52.4	7.3	615	US-08-998-416-186	Sequence 186, App
11	51	7.1	19124	US-08-487-826B-13	Sequence 13, Appl
12	50.2	7.0	678	US-07-991-867B-23	Sequence 23, Appl
13	50.2	7.0	678	US-08-107-755A-23	Sequence 23, Appl
14	50.2	7.0	678	US-08-544-332-23	Sequence 23, Appl
15	50.2	7.0	678	US-09-370-861A-23	Sequence 23, Appl
16	50.2	7.0	1788	US-09-601-198-101	Sequence 101, App
17	50.2	7.0	6124	US-08-213-419B-3	Sequence 3, Appl
18	50.2	7.0	6768	US-08-107-755A-1	Sequence 1, Appl
19	50.2	7.0	8457	US-07-991-867B-1	Sequence 1, Appl
20	50.2	7.0	8457	US-08-544-332-1	Sequence 1, Appl
21	50.2	7.0	8457	US-09-370-861A-1	Sequence 1, Appl
22	50	7.0	3232	US-09-333-214-3	Sequence 3, Appl
23	49.8	7.0	6243	US-09-056-075-1	Sequence 1, Appl
24	49.8	7.0	20674	US-09-641-638-651	Sequence 651, App
25	49.2	6.9	1692	US-09-601-198-63	Sequence 63, Appl
26	48.8	6.8	3234	US-08-286-325A-7	Sequence 7, Appl
27	48.6	6.8	615	US-08-998-416-186	Sequence 186, App

28	47.8	6.7	6124	US-08-213-419B-3	Sequence 3, Appl
29	47.6	6.7	711	US-09-601-198-111	Sequence 111, App
30	47.4	6.6	5340	US-09-627-122-21	Sequence 21, Appl
31	47.4	6.6	6152	US-08-973-462-1	Sequence 1, Appl
32	46.8	6.6	1511	US-07-991-867B-8	Sequence 8, Appl
33	46.8	6.6	1511	US-08-107-755A-8	Sequence 8, Appl
34	46.8	6.6	1511	US-08-544-332-8	Sequence 8, Appl
35	46.8	6.6	1511	US-09-370-861A-8	Sequence 8, Appl
36	46.8	6.6	2435	US-09-306-593-1	Sequence 1, Appl
37	46.2	6.5	827	US-08-998-416-535	Sequence 535, App
38	46	6.4	240	US-08-628-417-6	Sequence 6, Appl
39	46	6.4	658	US-08-998-416-595	Sequence 595, App
40	46	6.4	5852	US-07-867-106-2	Sequence 2, Appl
41	45.8	6.4	10640	US-09-417-485D-5	Sequence 5, Appl
42	45.6	6.4	636	US-08-998-416-1137	Sequence 1137, Ap
43	45.6	6.4	1474	US-08-821-994-64	Sequence 64, Appl
44	45.6	6.4	2435	US-09-306-593-1	Sequence 1, Appl
45	45.2	6.3	945	US-09-601-198-177	Sequence 177, App

## ALIGNMENTS

RESULT 1  
US-08-896-164-19  
Sequence 19, Application US/08896164  
Patent No. 6218521  
GENERAL INFORMATION:  
APPLICANT: OBATA, Yutichi  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED  
TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR  
TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/896,164  
FILING DATE: July 17, 1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6218521man D. Hanson  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 714 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-896-164-19

Query Match 99.4%, Score 710; DB 3; Length 714;  
Best Local Similarity 100.0%; Pred. No. 3.2e-132;  
Matches 714; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGAGAAAAGTATCTTCTATTAAACATCTTCTCAAGACGATATTTATCC 60  
DB 1 CGCGAGAAAAGTATCTTCTATTAAACATCTTCTCAAGACGATATTTATCC 60  
QY 61 TATATTCACGAGATTTTAAGAAATAACATTAGTATTGAAGAAACTGAGAAAAAGATTA 120

```

Db 61 TATATCTGCTGAATTTTAAAGAAATPACATTTAGTATTTAAGAAAACTAGAGAAAAAGATATA 120
Qy 121 ATGCAGATTAATTAACCTTCATGAAAAAGAAAAATTTATACAAAGACCTGAGACGTTAT 180
Db 121 ATGCAGATTAATTAACCTTCATGAAAAAGAAAAATTTATACAAAGACCTGAGACGTTAT 180
Qy 181 AAATGAATGAGATTAATTAATTTGAAAACTGCATCGAAAGCAACTTATTTGTTCAAT 240
Db 181 AAATGAATGAGATTAATTAATTTGAAAACTGCATCGAAAGCAACTTATTTGTTCAAT 240
Qy 241 ATNCTTAATGATGCTGTTTATGACTTAATACATGATTTTTCAGAGAGAAACCATGTT 300
Db 241 ATNCTTAATGATGCTGTTTATGACTTAATACATGATTTTTCAGAGAGAAACCATGTT 300
Qy 301 AAAAATATTTTATTTTAAATAATTAAGCTGTGTTCAGAGCTGTCATATTTCTTTTAT 360
Db 301 AAAAATATTTTATTTTAAATAATTAAGCTGTGTTCAGAGCTGTCATATTTCTTTTAT 360
Qy 361 TTGATTTGGGAAANAAATACGCTTCTGATAGCATGAAATGCAAAATTTTATGATTTTA 420
Db 361 TTGATTTGGGAAANAAATACGCTTCTGATAGCATGAAATGCAAAATTTTATGATTTTA 420
Qy 421 ATCTCACTAATTTTAAANACTATGAGAAATGATTAATGACATGACAGCAACACTA 480
Db 421 ATCTCACTAATTTTAAANACTATGAGAAATGATTAATGACATGACAGCAACACTA 480
Qy 481 ATTACTGCGCAGCTGTGGCATTGTGTTCTTACTTACTTCTCCCAAGGAAAACTCTTA 540
Db 481 ATTACTGCGCAGCTGTGGCATTGTGTTCTTACTTACTTCTCCCAAGGAAAACTCTTA 540
Qy 541 AATGAATCTTGAGAGAAATATCCTTAATATCTGTTGTAAGCAAAACAAAGCTTTT 600
Db 541 AATGAATCTTGAGAGAAATATCCTTAATATCTGTTGTAAGCAAAACAAAGCTTTT 600
Qy 601 TGTTTACATAGTCTTTGGGATTTTACTGTTCTTAATTTTATTCGAAACTCAATTTTAC 660
Db 601 TGTTTACATAGTCTTTGGGATTTTACTGTTCTTAATTTTATTCGAAACTCAATTTTAC 660
Qy 661 CCCAGACCATTAATACCATTAATCTGTTGTTGCAACAGTGTGTTGCCAATTC 714
Db 661 CCCAGACCATTAATACCATTAATCTGTTGTTGCAACAGTGTGTTGCCAATTC 714

```

## RESULT 2

```

US-08-896-164-20
; Sequence 20, Application US/08896164
; Patent No. 6218521
; GENERAL INFORMATION:
; APPLICANT: OBATA, Yuichi
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/896,164
; FILING DATE: July 17, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6218521man D. Hanson
; REGISTRATION NUMBER: 30,946

```

```

; REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-896-164-20

```

```

Query Match 85.7%; Score 611.8; DB 3; Length 687;
Best Local Similarity 97.1%; Pred. No. 8e-113;
Matches 671; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

```

```

Qy 15 ATTTTAATTTTCTATTAACATTTCTTCTCAAGCATTAATTTATCTATCTCACTGAA 74
Db 1 ATTTTAATTTTCTATTAACATTTCTTCTCAAGCATTAATTTATCTATCTCACTGAA 60
Qy 75 TTTTAAAGAAATTAACATTTAGTTTAAAGAAACTAGAGAAAAAGATTAATGAGATTAATTA 134
Db 61 TTTTAAAGAAATTAACATTTAGTTTAAAGAAACTAGAGAAAAAGATTAATGAGATTAATTA 120
Qy 135 ACTTACATGAAAGAAAGAAATTTATACAAAGACCTGAGACGTTATAATTTGAATGAGA 194
Db 121 ACTTACATGAAAGAAAGAAATTTATACAAAGACCTGAGACGTTATAATTTGAATGAGA 180
Qy 195 TTATATTTTGAAGAACTGACATCTGAAAGCAACTTATTTGTTCAATTAATNCTTAATGATG 254
Db 181 TTATATTTTGAAGAACTGACATCTGAAAGCAACTTATTTGTTCAATTAATNCTTAATGATG 240
Qy 255 TGTTTTATGACTTAATACCTGATTTTTCAGAGCAATTTTCTTTATTTGATTTGGGAAANA 314
Db 241 TGTTTTATGACTTAATACCTGATTTTTCAGAGCAATTTTCTTTATTTGATTTGGGAAANA 300
Qy 315 TTTTAAAGAAATTAAGCTGTGTTCAGAGCTGATCATATTTCTTTATTTGATTTGGGAAANA 374
Db 301 TTTTAAAGAAATTAAGCTGTGTTCAGAGCTGATCATATTTCTTTATTTGATTTGGGAAANA 360
Qy 375 AAATACGTTTCTGATAGCATGAAAGCAAAATTTTATGATTTTAACTGACATATTTT 434
Db 361 AAATACGTTTCTGATAGCATGAAAGCAAAATTTTATGATTTTAACTGACATATTTT 420
Qy 435 AANAACCTTAAGAAATTTGATTAATGACATGACAGCAACACTAATTTACTGCGCAGCT 494
Db 421 AAGAACCTTAAGAAATTTGATTAATGACATGACAGCAACACTAATTTACTGCGCAGCT 480
Qy 495 GTTGGCATTTGTTCTTACTTACTTACTTCTCCCAAGGAAAACTCTTAATTTGAATCTTTCAG 554
Db 481 GTTGGCATTTGTTCTTACTTACTTACTTCTCCCAAGGAAAACTCTTAATTTGAATCTTTCAG 539
Qy 555 CAGATAATTCCTTAATTAATTAATTTGTAAGCAAAACAAAGCTTTTGTGTTTACATAGTTC 614
Db 540 CNGATATA-CCTTAATTAATTAATTTGTAAGCAAAACAAAGCTTTTGTGTTTACATAGTTC 597
Qy 615 TTTGGGATTTTACTGTTCTTAATTTTATTCGAAAGCAATTTTACCACCAACATATTT 674
Db 598 TTT-GGATTTTACTGTTCTTAATTTTATTCGAAAGCAATTTTACCACCAACATATTT 656
Qy 675 ACCATA-TTAACTTGTGTTGACAGAGTGT 704
Db 657 ACCCTATTTAACTTTGTTATGACAGAGTGT 687

```

## RESULT 3

```

US-08-896-164-22
; Sequence 22, Application US/08896164
; Patent No. 6218521
; GENERAL INFORMATION:
; APPLICANT: OBATA, Yuichi
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR

```

```

; TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/896.164
; FILING DATE: July 17, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6218521man D. Hanson
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 688-3884
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 843 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-896-164-22
;
; Query Match      61.1%; Score 436.2; DB 3; Length 843;
; Best Local Similarity 73.1%; Pred. No. 4e-78;
; Matches 516; Conservative 0; Mismatches 187; Indels 3; Gaps 3;
;
; Oy 2 GCCAGAAAAGTATTTTATTTTCTATTAACATCTTCCAAAGCATTTATTTATCT 61
; Db 2 GCCAAAAAATTTATTTAAATTTCTATTAACNTCTCCCAAAACATTTATTTACCT 61
;
; Oy 62 ATATCTCATGTAATTTTAAAGAAATACATTTAGTATTAGAAAAAATAGAAAAATATAA 121
; Db 62 ATNNCNCNGAANTTTNAAAAAATACCTTTNNNTTTAAAAAACCTNGAAAAAATAAT 121
;
; Oy 122 TGCAGATTAATTAACCTTACATGAAAAAGAAAAATTAACAAGAGACTGAGAACGTTATA 181
; Db 122 NGCAATATNTTAACCTTCTGAAAAANGAAATTTNTACAAAGAGACGAAACNTTNTA 181
;
; Oy 182 AATTGAATGAGATTATTAATTTGAAAAAGTGCATCTGAAAAAGCAAACTTTATTTGTTCAATTA 241
; Db 182 ATTNGAANTTAAATTAATTAATTTNGAANGCGCAGCAGCAAACTTAAATGTTCAATTA 241
;
; Oy 242 TNCCTTAATGATGCTTTTATGATGATAATCACTGATTTTTCAGAAAGAAACCCATGTTA 301
; Db 242 TCTTAAGAAGGAGGNTTTNANNACTATATNCCNGATTTTCCAAATNGAANCCNNNTTA 301
;
; Oy 302 AAAATATTTTATTTTAAAAAATTAACCTGTGTTCAGCTCTGATGATATTTCTTTATTT 361
; Db 302 AAANTNTTTTATTTTAAAAAATTAACCCGNTTCCAAACCCGATGATCANATTTCTTTATTT 361
;
; Oy 362 TGAATTTGGGAANAATAATCTGTTTGTGATGATGATAAATGCAAAATTTTATGATTTTAA 421
; Db 362 GGATTTGGGAAAAAATAATNCNGTTCCNNATTAACCCNNAANNCAAAATTTTAAATTTTAA 421
;
; Oy 422 TCTCATATTTTAAANAACTATTTAGAAATGATTAATGACATGAAGTGCACAACTATA 481
; Db 422 CCCCTTANTTTTAAAAAATCTATNGAANAATNGATTAANGACTTGAATTCG-CAACCTTAN 480
;
; Oy 482 TTAAGGCGCAGCTGTGTCATTTGCTTTCTTAATTTTCCCAAGGAAAAAATCTTTAA 541
; Db 481 TTNCGGCCACACNGTGGGCTGNTGTTTCTTACTTANTTCCCCCAA-GGAAANNCCTTAA 539

```

```

; Oy 542 ATGATCTTCAGCAGATATACCTTAATAATACCTTTGTAGACAAAACAAAGCTTTT 601
; Db 540 NCGAANTCNCNCAAAATACCTTAANTATCTTGTAGAACAAACAAACCTTTT 599
;
; Oy 602 GTTTACATAGTCTTTGGGATTTTACTGTTCTTAATTTATTTCTGAAACTAATTTTACC 661
; Db 600 GTTTACNTA-NTCCTTGGGATTTTAAACGGGTCGCCCAATTTNATCCNGAACCAANTTTCCC 658
;
; Oy 662 CCAGACCAATAATTCATATTTAATTTGTTTNGCAGAGTTGTTGC 707
; Db 659 CCNACCATANTTACATTTTACCTGTGTAAGCGCAGTNGTTGC 704
;
; RESULT 4
; US-09-313-294A-1553/C
; Sequence 1553, Application US/0913294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laljudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313.294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 1553
; LENGTH: 173
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700551132H1
;
; US-09-313-294A-1553
;
; Query Match      23.2%; Score 165.6; DB 4; Length 173;
; Best Local Similarity 97.7%; Pred. No. 9.7e-25;
; Matches 168; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; Oy 8 AAAAGTATTTTAAATTTCTATTAACATTTCTTCAAGGATTTATTTATCTATATCT 67
; Db 172 AAAAGTATTTTAAATTTCTATTAACATTTCTTCAAGGATTTATTTATCTATATCT 113
;
; Oy 68 CACTGAATTTTAAAGAAATTAACATTTAGTTTGAAGAAAACTAGAAAAAAGATTAATGCAGA 127
; Db 112 CACTGAATTTTAAAGAAATTAACATTTAGTTTGAAGAAAACTAGAAAAAAGATTAATGCAGA 53
;
; Oy 128 TAATTAACCTTACATGAAAAAGAAAAATTAATTAACAAAGAGCTGAGAACGTTA 179
; Db 52 TAATTAACCTTACATGAAAAAGAAAAATTAATTAACAAAGAGCTGAGAACGTTA 1
;
; RESULT 5
; US-07-867-106-2/C
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r1s
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match      9.3%; Score 66.4; DB 1; Length 5852;
Best Local Similarity 46.8%; Pred. No. 5.3e-05;
Matches 240; Conservative 0; Mismatches 269; Indels 4; Gaps 1;

QY 16 TTTTAAATTTTCATTAACATCTCTCGAAGCATTTATTTATCCTATATCTCATGTAAT 75
DB 5751 TTTGTTTTTTTTTAATGATATGTTGTTGTTGTTTCTTAAATTTCTATTTTAA 5692
QY 76 TTTAAGAAATACATTAAGTATAGAAAGAACTAGAAAGAAAGATAATGACATTAATTA 135
DB 5691 TTTTAAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 5632
QY 136 CTATACATGAAAGAAATTAATTAACAAAGACTGAGAACGTTATTAATGAAATGAGAT 195
DB 5631 ATTAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 5572
QY 196 TATATTTGAAAGAACTGCAATCGAAGCAACCTTTTGTTCATTAATCTTAATGATGCT 255
DB 5571 TAAAAAAGTAAAGGGTTTTTTTTTAAATTAATGATGATTTTAAATTTAAATCATTTGA 5512
QY 256 GTTTTATGACTAATACACTGATTTTTCAGAGAGAAACCCATGTTAAATATTTTAAAT 315
DB 5511 CGAGATTAAAAATCTTAAACATAAACAATTAATTTGATTTTCTTTTCTTTTCTTTT 5452
QY 316 TT---AAAATAGCCGTGCTGATCAATTTCTTTTAAATTTGATTGGGA 371
DB 5451 TTTTAAATAATCAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5392
QY 372 ANAAAAATCTGTTTGTAGATGATAATGCAAAATTTTGAATTTTAAATCTCATAT 431
DB 5391 TAAATTTAATATATTTTAAATTTTAAATTAATGATCTATGATCTATTAATTTCCATG 5332
QY 432 TTTAANAATCTATGGAATGATTAATGACATGAGAGCAACAACATAATTAATGAGCA 491
DB 5331 TTTTAAATTTTCTTTTAAAGATTTTCTTTTAAATTAATTAATTAATTAATTAATTA 5272
QY 492 GCTGTGCAATGCTGTTCTTACTAGTCTCC 524
DB 5271 AATGATATCTGATCTCTTTTCCCAAGTTTCC 5239
```

```
RESULT 6
US-09-641-638-651/c
Sequence 651, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bouguetelert, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET.051CP1
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
```

LOCATION: 17063..17554  
OTHER INFORMATION: exon 14  
NAME/KEY: m3ec feature  
LOCATION: 17555..20674  
OTHER INFORMATION: 3' regulatory region  
NAME/KEY: allele  
LOCATION: 1128  
OTHER INFORMATION: 10-508-191 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1182  
OTHER INFORMATION: 10-508-245 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1559  
OTHER INFORMATION: 10-509-284 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1570  
OTHER INFORMATION: 10-509-295 : deletion of C  
NAME/KEY: allele  
LOCATION: 1827  
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTT  
NAME/KEY: allele  
LOCATION: 2048  
OTHER INFORMATION: 10-511-62 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 2323  
OTHER INFORMATION: 10-511-337 : insertion of T  
NAME/KEY: allele  
LOCATION: 2341  
OTHER INFORMATION: 10-512-36 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 2623  
OTHER INFORMATION: 10-512-318 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2832  
OTHER INFORMATION: 10-513-250 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2844  
OTHER INFORMATION: 10-513-262 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 2934  
OTHER INFORMATION: 10-513-352 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2947  
OTHER INFORMATION: 10-513-365 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 3802  
OTHER INFORMATION: 12-206-81 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 4062  
OTHER INFORMATION: 10-343-231 : deletion of C  
NAME/KEY: allele  
LOCATION: 4088  
OTHER INFORMATION: 12-206-366 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4109  
OTHER INFORMATION: 10-343-278 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4170  
OTHER INFORMATION: 10-343-339 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 5903  
OTHER INFORMATION: 10-346-23 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6019  
OTHER INFORMATION: 10-346-141 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6141  
OTHER INFORMATION: 10-346-263 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6183  
OTHER INFORMATION: 10-346-305 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6338

OTHER INFORMATION: 10-347-74 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6375  
OTHER INFORMATION: 10-347-111 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6429  
OTHER INFORMATION: 10-347-165 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6467  
OTHER INFORMATION: 10-347-203 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6484  
OTHER INFORMATION: 10-347-220 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6534  
OTHER INFORMATION: 10-347-271 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 6611  
OTHER INFORMATION: 10-347-348 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 7668  
OTHER INFORMATION: 10-348-391 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8608  
OTHER INFORMATION: 10-349-47 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 8658  
OTHER INFORMATION: 10-349-97 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8703  
OTHER INFORMATION: 10-349-142 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 8777  
OTHER INFORMATION: 10-349-216 : deletion of CTG  
NAME/KEY: allele  
LOCATION: 8785  
OTHER INFORMATION: 10-349-224 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 8926  
OTHER INFORMATION: 10-349-368 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12171  
OTHER INFORMATION: 10-350-72 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12429  
OTHER INFORMATION: 10-350-332 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13341  
OTHER INFORMATION: 10-507-170 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 13492  
OTHER INFORMATION: 10-507-321 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 13524  
OTHER INFORMATION: 10-507-353 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13535

Query Match 8.6%; Score 61.2; DB 4; Length 20674;  
Best Local Similarity 47.6%; Pred. No. 0.00064;  
Matches 207; Conservative 0; Mismatches 226; Indels 2; Gaps 1;

QY 9 AAAGTTATTTAAATTTCTATTAACAATCTTCTCAAGCATTTATTTATCCTATATCTC 68  
DB 11513 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11454  
QY 69 ACTGAATTTAAGAAATTAACATTTAGTATTAGAAAACAGAAAAGATAAATGACAGT 128  
DB 11453 A-AAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11396  
QY 129 AATTAACCTTACATGAAAGAAAGAAATTAATTAACAAGAGCTAGAACTTAATTAATGAA 188  
DB 11395 AATTATATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAA 11336

QY 189 ATGAGATTAATTTGAAAACTGTCGTAAGCAACTTATTTGTCATTAATCTTAA 248  
Db 11335 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11276  
QY 249 TGATGCTGTTTATGCTAATACAGTATTTTCAAGAAAGCAACCATGTTAAAAATAT 308  
Db 11275 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11216  
QY 309 TTTTATTTTAAATTAAGCCGTCGTCAGCTGTCATATTTCTTTATTTGATTG 368  
Db 11215 ATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTT 11156  
QY 369 GGAANAAATACGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTATCTCACT 428  
Db 11155 AATTATTAATTTAAATTAATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTA 11096  
QY 429 AATTTTAAATTAAT 443  
Db 11095 ATATTAATTAATTTT 11081

RESULT 7  
US-09-134-001C-150/C  
; Sequence 150, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 150  
; LENGTH: 927  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-150

Query Match 8.0%; Score 57.4; DB 4; Length 927;  
Best Local Similarity 49.7%; Pred. No. 0.0027;  
Matches 171; Conservative 0; Mismatches 169; Indels 4; Gaps 1;

QY 214 TCTGAAGCAAACTTATTTGTCATTAATTTAAATGATGCTTTTATGACTAATACAC 273  
Db 606 TATAATAGACTATTTTCTTATTTCTGTTTAAAGTTGCTTTTGAATTAATGATC 547  
QY 274 TGATTTTTCAGAGAAAGCAACCATGTTAAAAATTTTAAATTAATTAATTAATTAATTA 333  
Db 546 TAAATTTTTCATTAAGATTAATTTGCAAAATAGATCTTTAATGTAATTAATTTCCAGATAT 487  
QY 334 TCAAGCTGATCATATTTCTTTATTTTGAATTTGGGAAAAATACG-----GTTTCTGA 389  
Db 486 AAAAGCAATTCATTTCTTATTTCTTACATGTTCAATTAATTTAAATTAATTAATTAATTA 427  
QY 390 TAGCATGAATGCAAAATTTTATTTTATTTTATCTCAATTTTAAATTAATTAATTAATTA 449  
Db 426 TTTTAAATTAATTTGAATTTTGGATTTTCTCTTATTAATTCATTAATTTAGTAA 367  
QY 450 ATTGATTAATGACATGAAGTGCACAACATTAATTAATGAGCACTGTTGGCATTTGTTT 509  
Db 366 ATGACATGACATGTTGTTTCAATTTGCTCCATATGTTAATTAATTAATTTTATTTATTTGTTCT 307  
QY 510 CTTAGTTAGTCTCCCAAGGAAACCTGTAATTAATTAATTTCA 553  
Db 306 TACATTAATGTTCTGCTCATGTAATTAATTTGTAATTTCTTAA 263

RESULT 8  
US-08-998-416-288/C  
; Sequence 288, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jürgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reibschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 288:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PA61241RP  
US-08-998-416-288

Query Match 7.7%; Score 55.2; DB 3; Length 837;  
Best Local Similarity 46.4%; Pred. No. 0.0072;  
Matches 212; Conservative 0; Mismatches 241; Indels 4; Gaps 1;

QY 5 AGAAAAAGTTATTTTAAATTTTCTATTAACATTTTCTCAAGCATTTATTTATCTATA 64  
Db 579 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 520  
QY 65 TCTACTGAATTTTAAAGAAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 124  
Db 519 ATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 460  
QY 125 AGATTAATTAATTTACATGAAAAAGAAAAATTAATTAACAAAGCATGTAATTAAT 184  
Db 459 TAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 400  
QY 185 TGAATGAGATTAATTTTGAAGAACTGCATCTGAAGCAAACTTATTTGTTCAATTAAT 244  
Db 399 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 340  
QY 245 TTAATGATGTTGTTTATGACTAATACATGATTTTTCAGAAAGAAACCATGTTAA 304

Db 339 ATTATTTTAATTAACAAATTAATAA-----TAATATATTAATTAATGATATCTATTTAT 284  
Qy 305 ATATTTTATTTTAAATAAAGCTGTGTCAGCTGATCATATTTCTTTATTTTGA 364  
Db 283 AATTTATTAAGAAATAATATATCTAATATATTTTAAATCAATTTTAAATTTGA 224  
Qy 365 TTGGGAANNAATACGTCTTCTGATAGCATGAATGCAAAATTTTGATTTTAACT 424  
Db 223 CATGACTAAATAGTATTCATATTAATAATATTTTAAATATATTAATTAAT 164  
Qy 425 CACTAATTTTAAANACTATGAGAAATGATTAATGA 461  
Db 163 GATGAATTAAGTAAATTAATTAATTAATTAATTA 127

RESULT 9  
US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 7.4%; Score 53; DB 2; Length 19124;  
Best Local Similarity 45.7%; Pred. No. 0.026;  
Matches 215; Conservative 0; Mismatches 253; Indels 2; Gaps 1;

Qy 6 GAAAAAGTTATTTATTTCTATTAACATCTCTCAAGACATTAATTTATCTATAT 65  
Db 15367 GATGACTGATGATGCTTTATATATATATATATATATATATATTAATTAAGAA 15426  
Qy 66 CTCAGTGAATTTTAAGAAATAACATTAGATTAGAAAAAACTAGAAAAAAGTAAATGCA 125

Db 15427 AATTAATAAACAATTTATTAATTAATGAAAAAAGAAAAATGAAATATTAATTAATTA 15486  
Qy 126 GATAATTAACCTACATGAAAAAGAAATTAATAACAAGACCTGACATTAATTT 185  
Db 15487 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15546  
Qy 186 GAATGAGATTAATTAATTTGAAAACTGCATCGAAGCAACTTAATGTTAATTAATCT 245  
Db 15547 ATAAATAAATAAATAAATTTGATA--GAATTAATAAATAAATAAATAAATAAATAA 15604  
Qy 246 TAATGATGCTGTTTATGACTAATATACATGATTTTTCAGAAGAAACCCATGTTAAAA 305  
Db 15605 AAAAAAAAAATTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15664  
Qy 306 TATTTTATTTTAAATAAATAAGCCTGTGTCAGCTCGATCATATTTCTTTATTTGAT 365  
Db 15665 AAAAAAAAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 15724  
Qy 366 TTGGGAANNAATACGTCTTCTGATAGCATGAATGCAAAATTTTATTAATCTC 425  
Db 15725 ATATTAATAATTAATAATATATATCATTAATAAATAAATAAATAAATAAATAA 15784  
Qy 426 ACTAATTTTAAANACTATGAGAAATGATTAATGACATGAAGTGCACAA 475  
Db 15785 AATATATCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 15834

RESULT 10  
US-08-998-416-186  
Sequence 186, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippseu, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Rebeschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYP11  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG107ARP  
US-08-998-416-186

Query Match 7.3%; Score 52.4; DB 3; Length 615;  
Best Local Similarity 47.0%; Pred. No. 0.025;  
Matches 225; Conservative 0; Mismatches 249; Indels 5; Gaps 2;

QY 5 AGAAAAAGTATTTTAAATTTCTATTAACATCTTCTCAAGCAATTTATTCCTATA 64  
DB 26 ATAAAGATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAAT 85  
QY 65 TCTCACTGATTTTAAAGAAATACATAGTATTAAGAAAACAGAAAAAATTAATGC 124  
DB 86 ATTATCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 141  
QY 125 AGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 184  
DB 142 TATATAATTTACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 201  
QY 185 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 244  
DB 202 TATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 260  
QY 245 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 304  
DB 261 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 320  
QY 305 ATATTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 364  
DB 321 AATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 380  
QY 365 TTGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 424  
DB 381 AATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 440  
QY 425 CACTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 483  
DB 441 AACTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 499

## RESULT 11

US-08-487-826B-13/C  
Sequence 13, Application US/08487826B

GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chien, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhuan  
APPLICANT: Williams, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 7.1%; Score 51; DB 2; Length 19124;  
Best Local Similarity 45.8%; Pred. No. 0.065;  
Matches 171; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 9 AAATTTATTTTAAATTTCTATTAACATCTTCTCAAGCAATTTATTCCTATATCTC 68  
DB 18383 ATATTTTATTTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTAT 18324  
QY 69 ACTGATTTTAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 128  
DB 18323 ATGTAATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 18264  
QY 129 AATTAATTTCTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 188  
DB 18263 AGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18204  
QY 189 ATGATTAATTAATTTGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 248  
DB 18203 ATTAATTAATTTATTTTAAAGAAATTAATTAATTAATTAATTAATTAATTAAT 18144  
QY 249 TGATGCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 308  
DB 18143 TGATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 18084  
QY 309 TTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 368  
DB 18083 ACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18024  
QY 369 GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 381  
DB 18023 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18001

## RESULT 12

US-07-991-867B-23  
Sequence 23, Application US/07991867B

GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanichk  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS



SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/991,867B  
FILING DATE: 12-DEC-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UFI14.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 678 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-07-991-867B-23

Query Match 7.0%; Score 50.2; DB 1; Length 678;  
Best Local Similarity 46.1%; Pred. No. 0.069; Indels 0; Gaps 0;

Matches 166; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 5 AGAAAAAGTTATTTTCTATTAACATTTCTTCAAGCATTTTATCTCTATA 64  
DB 31 ATAATAATTTATTTATTTGATTAATTTATTTATTTATTTATTTATTTA 90  
QY 65 TCTCAGCAATTTTAAAGAAATTAACATTGATTTAGAAAACTAGAAAAAGATAATGC 124  
DB 91 GATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAC 150  
QY 125 AGATAATTAACCTTACATGAAGAAAGAAATTAATTAACAAAGACTGAGAACTTAAT 184  
DB 151 AATAATAATTAATTTAATTTATTAAGATTAATTTATTTATTTATTTATTTATTTAAC 210  
QY 185 TGAATGAGATTATTAATTTGAAAACTGCATCTGAAGCAAACTTTATTTGTTCAATTATNC 244  
DB 211 ATAATAATTAATTTATTTATTTAGTAACACAGCTAATTAATTTATTTATTTATTTATTTAAT 270  
QY 245 TTAATGATGCTTTTATTTAGCTAATTAACCTGATTTTCAAGAGAAACCAATTTAA 304  
DB 271 GTAATAATTAATTTATTTATTTAGTAACACAGCTAATTAATTTATTTATTTATTTAAT 330  
QY 305 ATATTTTATTTTAAATAAGCCGCTGCTGATCAAGCTGATCAATTTCTTTATTTTGA 364  
DB 331 ATATTTATTTAGATTAAAGAAACATTAATTAATTAATTAATTAATTTATTTATTTAGA 390

RESULT 13  
US-08-107-755A-23  
Sequence 23, Application US/08107755A  
Patent No. 5721352  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Guidi, Michael E.  
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville

STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,755A  
FILING DATE: 19-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,658  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UFI14.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 678 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-107-755A-23

Query Match 7.0%; Score 50.2; DB 1; Length 678;  
Best Local Similarity 46.1%; Pred. No. 0.069; Indels 0; Gaps 0;

Matches 166; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 5 AGAAAAAGTTATTTTCTATTAACATTTCTTCAAGCATTTTATCTCTATA 64  
DB 31 ATAATAATTTATTTATTTGATTAATTTATTTATTTATTTATTTATTTA 90  
QY 65 TCTCAGCAATTTTAAAGAAATTAACATTGATTTAGAAAACTAGAAAAAGATAATGC 124  
DB 91 GATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAC 150  
QY 125 AGATAATTAACCTTACATGAAGAAAGAAATTAATTAACAAAGACTGAGAACTTAAT 184  
DB 151 AATAATAATTAATTTAATTTATTAAGATTAATTTATTTATTTATTTATTTATTTAAC 210  
QY 185 TGAATGAGATTATTAATTTGAAAACTGCATCTGAAGCAAACTTTATTTGTTCAATTATNC 244  
DB 211 ATAATAATTAATTTATTTATTTAGTAACACAGCTAATTAATTTATTTATTTATTTAAT 270  
QY 245 TTAATGATGCTTTTATTTAGCTAATTAACCTGATTTTCAAGAGAAACCAATTTAA 304  
DB 271 GTAATAATTAATTTATTTATTTAGTAACACAGCTAATTAATTTATTTATTTATTTAAT 330  
QY 305 ATATTTTATTTTAAATAAGCCGCTGCTGATCAAGCTGATCAATTTCTTTATTTTGA 364  
DB 331 ATATTTATTTAGATTAAAGAAACATTAATTAATTAATTAATTAATTTATTTATTTAGA 390

RESULT 14  
US-08-544-332-23  
Sequence 23, Application US/08544332  
Patent No. 5935777  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Guidi, Michael E.  
TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 77



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:02:42 ; Search time 2782.58 Seconds

(without alignments)  
10100.295 Million cell updates/sec

Title: US-09-835-992A-20

Perfect score: 687

Sequence: 1 attctaattctctcttaaac.....cttgctatgcacagctgctc 687

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba:\*

2: gb\_hcg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pac:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sbs:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: gb\_da:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_ov:\*

22: em\_ov:\*

23: em\_pac:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sbs:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pln:\*

35: em\_hcg\_rtd:\*

36: em\_hcg\_mam:\*

37: em\_hcg\_vrc:\*

38: em\_sy:\*

39: em\_hcgo\_hum:\*

40: em\_hcgo\_mus:\*

41: em\_hcgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	683	99.4	687	6	AR146580
2	683	99.4	687	6	BD079829
3	638.4	92.9	1052	9	HSSCPX16
4	638.4	92.9	2572	2	HUMSCP2A
5	638.4	92.9	122176	2	AL358233
6	638.4	92.9	175046	2	AC0272728
7	638.4	92.9	193774	2	AL445183
8	636.8	92.7	1439	9	BC005911
9	611.8	89.1	714	6	AR146579
10	611.8	89.1	714	6	BD079828
11	455.6	66.3	1229	9	S52450
12	440.4	64.1	1219	9	HUMSTPAA
13	411.8	59.9	843	6	AR146582
14	411.8	59.9	843	6	BD079831
15	296.4	43.1	1500	9	HUMSCP2B
16	216.4	31.5	2661	4	AF051897
17	180	26.2	145564	2	AC134794
18	179	26.1	2626	10	BC018384
19	179	26.1	248677	2	AL844206
20	177.4	25.8	2599	10	RAT60KDA
21	168.8	24.6	2571	10	RATSCPXA
22	161	23.4	173	6	AR246194
23	154.6	22.5	904	10	S80339
24	126.2	18.4	263	6	A74403
25	126.2	18.4	263	6	A77382
26	85.8	12.5	2152	6	AX306091
27	85.8	12.5	2152	10	MUSSCP
28	81.8	11.9	95477	9	AC007076
29	77	11.2	168799	9	AC009531
30	75.8	11.0	125623	3	AC115599
31	74.8	10.9	113202	9	HSU37823
32	74.6	10.9	147956	2	AC137839
33	74.4	10.8	152878	3	CEY18D10A
34	73.8	10.7	149597	2	AC034271
35	73.8	10.7	157749	2	AC025535
36	73.2	10.7	251762	3	AE014851
37	73.2	10.7	310779	2	AC005140
38	72.6	10.6	1192	9	HSAL33759
39	72.6	10.6	158548	3	PFMAL3P2
40	72.2	10.5	5850	3	DDIDP2
41	72.2	10.5	182870	3	AC116960
42	72	10.5	151259	2	AC138822
43	72	10.5	160701	9	AC108106
44	72	10.5	172404	9	AC138948
45	72	10.5	181154	2	AC138842

#### ALIGNMENTS

RESULT 1

AR146580

LOCUS

DEFINITION Sequence 20 from patent US 6218521.

ACCESSION AR146580

VERSION AR146580.1 GI:15109769

KEYWORDS

SOURCE

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 687)

AUTHORS Obata, Y.

TITLE Isolated nucleic acid molecules associated with gastric cancer and methods for diagnosis and treating gastric cancer

JOURNAL Patent: US 6218521-A 20 17-APR-2001;

FEATURES Location/Qualifiers  
 source 1..687  
 /organism="unknown"  
 BASE COUNT 242 a 98 c 85 g 258 t 4 others  
 ORIGIN

Query Match 99.4%; Score 683; DB 6; Length 687;  
 Best Local Similarity 100.0%; Pred. No. 1,4e-91;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTAATTTTCATTAACAATCTCTTCAAGCATTAATTTATCCATATCTCACTGAA 60  
 DB 1 ATTTAATTTTCATTAACAATCTCTTCAAGCATTAATTTATCCATATCTCACTGAA 60  
 QY 61 TTTTAAANAATTAACCTTGATTAAGAAAACCTAGAAAAGATTAAGCAGATATTAA 120  
 DB 61 TTTTAAANAATTAACCTTGATTAAGAAAACCTAGAAAAGATTAAGCAGATATTAA 120  
 QY 121 ACTTACATGAAGAAAGAAATTTATPACAAGAGACTGAGAGCTTATTAATTGAATGAGA 180  
 DB 121 ACTTACATGAAGAAAGAAATTTATPACAAGAGACTGAGAGCTTATTAATTGAATGAGA 180  
 QY 181 TTATTAATTTGAAAACCTGCATCTGAAAAGCAACTTATTTGTCATTAATTTCTTAAATGATG 240  
 DB 181 TTATTAATTTGAAAACCTGCATCTGAAAAGCAACTTATTTGTCATTAATTTCTTAAATGATG 240  
 QY 241 TGTTTTATGACTTAATACCTGATTTTTCATTAAGAAACCAATGTTAAATATTTTAT 300  
 DB 241 TGTTTTATGACTTAATACCTGATTTTTCATTAAGAAACCAATGTTAAATATTTTAT 300  
 QY 301 TTTTAAANAATTAAGCTGTGTCAAGCTGATCATATTTCTTTTATTTGATTTGGGAAGA 360  
 DB 301 TTTTAAANAATTAAGCTGTGTCAAGCTGATCATATTTCTTTTATTTGATTTGGGAAGA 360  
 QY 361 AAATACGTGTTTCGATAGCATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420  
 DB 361 AAATACGTGTTTCGATAGCATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420  
 QY 421 AAGAACCTTTGGAAGAAATGATTAATGACATGAAGTGCACACATTAATTTACTGGCAGCT 480  
 DB 421 AAGAACCTTTGGAAGAAATGATTAATGACATGAAGTGCACACATTAATTTACTGGCAGCT 480  
 QY 481 GTTGGCATTTGTTTCTTACTTAAGTTCGCCAAGAAACTCTTAACTGAATCTTCAGC 540  
 DB 481 GTTGGCATTTGTTTCTTACTTAAGTTCGCCAAGAAACTCTTAACTGAATCTTCAGC 540  
 QY 541 NGAATTAACCTTAATATATCTTTGAGCAAAACAAACTTTTGTTTACATAGTCTTT 600  
 DB 541 NGAATTAACCTTAATATATCTTTGAGCAAAACAAACTTTTGTTTACATAGTCTTT 600  
 QY 601 GGATTTTACTGTTCTCTAATTTTATTTCTGAACTCATTTTCCCAAGCCATTAATCCC 660  
 DB 601 GGATTTTACTGTTCTCTAATTTTATTTCTGAACTCATTTTCCCAAGCCATTAATCCC 660  
 QY 661 TATTTAATTTGTTATGACAGAGTTGTT 687  
 DB 661 TATTTAATTTGTTATGACAGAGTTGTT 687

RESULT 2  
 BD079829 687 bp DNA linear PAT 27-AUG-2002  
 LOCUS  
 DEFINITION Cancer-associated nucleic acids and polypeptides.  
 ACCSSION BD079829  
 VERSION BD079829.1 GI:22625432  
 KEYWORDS JP 2001516009-A/495.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 687)  
 Old, L. J., Scanlan, M. J., Stockert, E., Gure, A., Chen, Y. T., Gout, I.,  
 Ogbare, M., Obara, Y., Pfreundschuh, M., Tureci, O. and Sahin, U.

TITLE  
 JOURNAL  
 COMMENT Cancer-associated nucleic acids and polypeptides  
 Patient: JP 2001516009-A 495 25-SEP-2001;  
 LUDWIG INSTITUTE FOR CANCER RESEARCH  
 OS Homo sapiens (human)  
 PN JP 2001516009-A/495  
 PD 25-SEP-2001  
 PR 15-JUL-1998 JP 2000503425  
 PR 17-JUL-1997 US 08/896164,10-OCT-1997 US 60/061599 PR  
 10-OCT-1997 US 60/061765,10-OCT-1997 US 08/948705 PR  
 11-OCT-1997 GB 5721697.2,22-JUN-1998 US 09/102322 PI  
 J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI  
 CHEN,  
 PI IVAN GOUT, MICHAEL O'HARE, YUICHI OBATA, MICHAEL Pfreundschuh, PI  
 OZLEM TURECI  
 PI UGUR SAHIN

FEATURES Location/Qualifiers  
 source 1..687  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 242 a 98 c 85 g 258 t 4 others  
 ORIGIN

Query Match 99.4%; Score 683; DB 6; Length 687;  
 Best Local Similarity 100.0%; Pred. No. 1,4e-91;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTAATTTTCATTAACAATCTCTTCAAGCATTAATTTATCCATATCTCACTGAA 60  
 DB 1 ATTTAATTTTCATTAACAATCTCTTCAAGCATTAATTTATCCATATCTCACTGAA 60  
 QY 61 TTTTAAANAATTAACCTTGATTAAGAAAACCTAGAAAAGATTAAGCAGATATTAA 120  
 DB 61 TTTTAAANAATTAACCTTGATTAAGAAAACCTAGAAAAGATTAAGCAGATATTAA 120  
 QY 121 ACTTACATGAAGAAAGAAATTTATPACAAGAGACTGAGAGCTTATTAATTGAATGAGA 180  
 DB 121 ACTTACATGAAGAAAGAAATTTATPACAAGAGACTGAGAGCTTATTAATTGAATGAGA 180  
 QY 181 TTATTAATTTGAAAACCTGCATCTGAAAAGCAACTTATTTGTCATTAATTTCTTAAATGATG 240  
 DB 181 TTATTAATTTGAAAACCTGCATCTGAAAAGCAACTTATTTGTCATTAATTTCTTAAATGATG 240  
 QY 241 TGTTTTATGACTTAATACCTGATTTTTCATTAAGAAACCAATGTTAAATATTTTAT 300  
 DB 241 TGTTTTATGACTTAATACCTGATTTTTCATTAAGAAACCAATGTTAAATATTTTAT 300  
 QY 301 TTTTAAANAATTAAGCTGTGTCAAGCTGATCATATTTCTTTTATTTGATTTGGGAAGA 360  
 DB 301 TTTTAAANAATTAAGCTGTGTCAAGCTGATCATATTTCTTTTATTTGATTTGGGAAGA 360  
 QY 361 AAATACGTGTTTCGATAGCATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420  
 DB 361 AAATACGTGTTTCGATAGCATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420  
 QY 421 AAGAACCTTTGGAAGAAATGATTAATGACATGAAGTGCACACATTAATTTACTGGCAGCT 480  
 DB 421 AAGAACCTTTGGAAGAAATGATTAATGACATGAAGTGCACACATTAATTTACTGGCAGCT 480  
 QY 481 GTTGGCATTTGTTTCTTACTTAAGTTCGCCAAGAAACTCTTAACTGAATCTTCAGC 540  
 DB 481 GTTGGCATTTGTTTCTTACTTAAGTTCGCCAAGAAACTCTTAACTGAATCTTCAGC 540  
 QY 541 NGAATTAACCTTAATATATCTTTGAGCAAAACAAACTTTTGTTTACATAGTCTTT 600

|||||  
Db 541 NGATAAACCCTAAATATATCTTTGTTAGCCAAACAACCTTTTGTGTACATAGTCTTT 600  
|||  
Qy 601 GGATTTTACTGTTCTCTAATTTTATTCGAAACCTCATTTTCCCAACCTAATTTACCC 660  
|||  
Db 601 GGATTTTACTGTTCTCTAATTTTATTCGAAACCTCATTTTCCCAACCTAATTTACCC 660  
|||  
Qy 661 TATTTAATCTTGTATGACAGTGTGT 687  
|||  
Db 661 TATTTAATCTTGTATGACAGTGTGT 687  
|||  
RESULT 3  
HSCSPT16/c 1052 bp DNA linear PRI 25-JAN-1997  
LOCUS Human sterol carrier protein-X/sterol carrier protein-2  
DEFINITION (SCP-X/SCP-2) gene, exon 16, and complete cds.  
ACCESSION U11313  
VERSION U11313.1  
KEYWORDS GI:532077  
SEGMENT 16 of 16  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1052)  
AUTHORS Ohba,T., Renner,H., Pfeifer,S.M., He,Z., Yamamoto,R., Holt,J.A.,  
Billheimer,U.T. and Straus,U.F. III.  
TITLE The structure of the human sterol carrier protein X/sterol carrier  
JOURNAL protein 2 gene (SCP2)  
MEDLINE Genomics 24 (2), 370-374 (1994)  
PUBMED 95213031  
REFERENCE 2 (bases 1 to 1052)  
AUTHORS Straus,U.F. III.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUN-1994) Jerome F. Straus III, Department of  
Obstetrics and Gynecology, Division of Reproductive Biology,  
University of Pennsylvania Medical Center, 778 Clinical Research  
Building, 422 Curie Boulevard, Philadelphia, PA 19104-6142, USA  
location/Qualifiers  
FEATURES  
source 1..1052  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p32"  
/cfeature\_type="1iver"  
gene join(U11297.1:1..3666,U11299.1:1..108,U11300.1:1..122,  
U11301.1:1..182,U11302.1:1..115,U11303.1:1..177,  
U11304.1:1..114,U11305.1:1..137,U11306.1:1..201,  
U11307.1:1..198,U11308.1:1..158,U11309.1:1..175,  
U11310.1:1..153,U11311.1:1..180,U11312.1:1..130,1..1034)  
/partial  
mRNA /gene="SCP-X/SCP-2"  
join(U11297.1:3452..3641,U11299.1:26..83,U11300.1:26..97,  
U11301.1:26..157,U11302.1:26..90,U11303.1:26..152,  
U11304.1:26..89,U11305.1:26..112,U11306.1:26..176,  
U11307.1:26..173,U11308.1:26..133,U11309.1:26..1700,  
U11310.1:26..128,U11311.1:26..155,U11312.1:26..105,  
26..1034)  
/gene="SCP-X/SCP-2"  
/product="sterol carrier protein-X/sterol carrier  
protein-2"  
CDS join(U11297.1:3573..3641,U11299.1:26..83,U11300.1:26..97,  
U11301.1:26..157,U11302.1:26..90,U11303.1:26..152,  
U11304.1:26..89,U11305.1:26..112,U11306.1:26..176,  
U11307.1:26..173,U11308.1:26..133,U11309.1:26..1700,  
U11310.1:26..128,U11311.1:26..155,U11312.1:26..105,  
26..121)  
/gene="SCP-X/SCP-2"  
/codon\_start=1  
/product="sterol carrier protein-X/sterol carrier

protein-2"  
protein\_id="AA041286.1"  
/db\_xref="GI:1773239"  
/translation="MSSSPWEPAAALRVFVGVGNTKFKVPGAENSDYDPLAEAGK  
KALADAI PYSADVQACVGFSDSTGQSAIYHSLDMTGIPIINVNNCAVSTALF  
MARLADIGVAECVLAIGFERKMSKSGISGIFSDTIPDRIKVDLLIKYGISAPVAP  
OMFYAGKEHMEKGTKEIHFPAKIGMNGHGSVNNPYSDQDEYSLBVMKSVFPF  
LTILOCCPTSDGAATAIASEAFVOKYGLDSKAVEILAOEMMDLPSFEEKSI1IKVY  
GPMSEKAAKCYEKSGLTPNDIDVIELHDCFSNELLYEALGLCEPGCATIYDRG  
DNYYGKRVINPSGGLISKHPLGATLACAEICWOLRBEAGRQVPGARVALIQHL  
DIGAVVVTLYKMGFEPAASSFRHQIEAVPTSASDGFANLVFKEIEKLEEGEQ  
FVKIGIGIFAFKCYDGPGEKATWVVDVKNKGSGLVPSDKDCITTMADSDSLALM  
TGKNNPOSFAFFQGLKITGNMGLAMKLQNLQPGNAKL"  
exon 26..1034  
/gene="SCP-X/SCP-2"  
/note="nucleotides 122-1027 are not translated"  
/number=16  
BASE COUNT 352 a 157 c 164 g 379 t  
ORIGIN  
Query Match 92.9%; Score 638.4; DB: 9; Length 1052;  
Best Local Similarity 98.1%; Pred. No. 4.6e-85;  
Matches 675; Conservative 0; Mismatches 10; Indels 3; Gaps 3;  
Qy 1 ATTTAATTTTCTTAAACATCTCTCAAGACATTTTATCTATATCTGACGAA 60  
|||  
Db 1014 ATTTAATTTTCTTAAACATCTCTCTCAAGACATTTTATCTATATCTGACGAA 955  
|||  
Qy 61 TTTTAAATAATTAACATTAGTATTAAGAAAACTAGAAAAAGATMAATGACATATTAA 120  
|||  
Db 954 TTTTAAATAATTAACATTAGTATTAAGAAAACTAGAAAAAGATMAATGACATATTAA 895  
|||  
Qy 121 ACTTACATGAAAAAGAAAAATTTAACAAGAGATGGAACGTATTAATTTGAATGGA 180  
|||  
Db 894 ACTTACATGAAAAAGAAAAATTTAACAAGAGATGGAACGTATTAATTTGAATGGA 835  
|||  
Qy 181 TTATATTTGAAAACTGATCTGAAAGCAAACTTATTTGTCATTTTCTTAATGATGG 240  
|||  
Db 834 TTATATTTGAAAACTGATCTGAAAGCAAACTTATTTGTCATTTTCTTAATGATGG 775  
|||  
Qy 241 TGTTTTATGACTATATACATGATTTTCAATTAAGAAACCAGTTAAATAATTTTAT 300  
|||  
Db 774 TGTTTTATGACTATATACATGATTTTCAATTAAGAAACCAGTTAAATAATTTTAT 715  
|||  
Qy 301 TTTTAAATAATTAACCTGTGTTCAAGCTCTGATCATTTTCTTTATTTTGAATGGGA 360  
|||  
Db 714 TTTTAAATAATTAACCTGTGTTCAAGCTCTGATCATTTTCTTTATTTTGAATGGGA 655  
|||  
Qy 361 AAATACGTTTCTGATGATGATGAAGCAAAATTTTGAATTTTGAATCTGCTAATTTT 420  
|||  
Db 654 AAATACGTTTCTGATGATGATGAAGCAAAATTTTGAATTTTGAATCTGCTAATTTT 595  
|||  
Qy 421 AAGAACTATTGAGAAATTTGATTAATGATGATGATGATGATGATGATGATGATGATG 480  
|||  
Db 594 AAGAACTATTGAGAAATTTGATTAATGATGATGATGATGATGATGATGATGATGATG 535  
|||  
Qy 481 GTTGCACTGTGTTTCTTACTTACTTCTCCAGAGAAAACTTTAACTGAATCTTACG 540  
|||  
Db 534 GTTGCACTGTGTTTCTTACTTACTTCTCCAGAGAAAACTTTAACTGAATCTTACG 475  
|||  
Qy 541 NGATAA-CCTTAATATCTGTTGTAGCCAAACAAA-CCTTTTGTATACATGTTCT 598  
|||  
Db 474 AGAATAATCCTTAATATCTGTTGTAGCCAAACAAA-CCTTTTGTATACATGTTCT 415  
|||  
Qy 599 TTGATTTTACTGTTCTTAAATTTTATCTGAAACCTCATTTTCCAGACCATTAATTAC 658  
|||  
Db 414 TTGATTTTACTGTTCTTAAATTTTATCTGAAACCTCATTAATTACCCAGACCATTAATTAC 355  
|||  
Qy 659 CTTATTTAATCTTGTATGACAGTGT 686  
|||  
Db 354 CATA-TTAATCTTGTATGACAGTGT 328  
|||

RESULT 4  
HUMSCP2A/LOCUS 2572 bp mRNA linear PRI 06-DEC-1993  
DEFINITION Human sterol carrier protein X/sterol carrier protein 2 mRNA,  
complete cds.  
ACCESSION M75883  
VERSION M75883.1 GI:432974  
KEYWORDS sterol carrier protein-2, sterol carrier protein X.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2572)  
AUTHORS He,Z., Yamamoto,R., Furch,E.E., Schantz,L.J., Naylor,S.L.,  
George,H., Billheimer,J.T. and Strauss,J.F. III.  
CDNAS encoding members of a family of proteins related to human  
sterol carrier protein 2 and assignment of the gene to human  
chromosome 1 p21---pter  
JOURNAL DNA Cell Biol. 10 (8), 559-569 (1991)  
MEDLINE 92029618  
PUBMED 1718316  
REFERENCE 2 (bases 1 to 2572)  
AUTHORS Vesa,J., Hellsten,E., Branoski,B.L., Emanuel,B.S., Billheimer,J.T.,  
Mead,S., Cowell,J.K., Strauss,J.F. III. and Peltonen,L.  
Assignment of sterol carrier protein X/sterol carrier protein 2 to  
1p32 and exclusion as the causative gene for infantile neuronal  
ceroid lipofusiosis  
JOURNAL Unpublished  
COMMENT On Dec 6, 1993 this sequence version replaced gi:410029.  
FEATURES  
source  
1..2572  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="1p32"  
/cissue\_type="liver"  
/clone\_id="Clontech HL-1001b/HL-1115b"  
1..2572  
/gene="SCP-X/SCP-2"  
22..1665  
/gene="SCP-X/SCP-2"  
/product="sterol carrier protein X/sterol carrier  
protein-2"  
/protein\_id="AA03557.1"  
/db\_xref="GI:432975"  
/translation="MSSSPMEPATLRVGVVGVGTRFKVKGAEHSRDYDLAEAGK  
KALADAOIPYAVDQACVGFSDTCGORAIYSLQMTGPIINNNNNCATGSTALF  
MAROLIOGVAEVCLALGFEKMSKSLGIKPSDRITIPDKHDLINKYGLSAHVAP  
CMRGYAKREHEKYEKCTIEHAKYGMKNGHNSVNNPYSPQDEVSIDLEWASKEVPD  
LTIIQCCPTSDGAATAILASERAFQYKGLQSKAVEIIAQMEDTLPSPFEKSLITIKV  
GFDMSKEAARCTCYKSGLTPTNDIVIELHDFSTNELITTEALGLCEGQCATLVLDG  
DNVYKGVINPISGGLISKHPLGATLAQACELCWLIRGAGRGVGAVALQHNIL  
GIGAVVTVLTKMFPASASFRTHQIEAVPTSSASDGLKLVFKIEIKLEEGGQ  
FVKIKGIFAFKVDGPGGKEATVVDVVDKSGSVLTPNSDKACCTTTMADSDPLAM  
TGKNPQSAFPGGLKITGNGLAMKQNLQPGNAKL"  
1932..1937  
/gene="SCP-X/SCP-2"  
/evidence=not\_experimental  
2213..2218  
/gene="SCP-X/SCP-2"  
/evidence=not\_experimental  
2258..2263  
/gene="SCP-X/SCP-2"  
/evidence=not\_experimental  
2341..2346  
/gene="SCP-X/SCP-2"  
/evidence=experimental  
2550..2555  
/gene="SCP-X/SCP-2"  
/evidence=experimental  
2572  
polya\_signal  
polya\_signal  
polya\_signal  
polya\_signal  
polya\_signal  
polya\_site

BASE COUNT 794 a 439 c 570 g 769 t  
ORIGIN  
Query Match 92.9%; Score 638.4; DB 9; Length 2572;  
Best Local Similarity 98.1%; Pred. No. 3.5e-85;  
Matches 675; Conservative 0; Mismatches 10; Indels 3; Gaps 3;  
/gene="SCP-X/SCP-2"  
/evidence=experimental  
1 ATTTAATTTTCTATTAACATTTCTTCAAGCATTTATTTATCTCATCTACGAA 60  
2558 ATTTAATTTTCTATTAACATTTCTTCAAGCATTTATTTATCTCATCTACGAA 2499  
61 TTTAATAAATAATTAACATTTATTAAGAAAGTAAAGAAAGTAAAGTAAAGTAA 120  
2498 TTTAAGAAATAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2439  
121 ACTTACATGAAAAAGAAAAATTAATTAAGAAAGTAAAGTAAAGTAAAGTAAAGT 180  
2438 ACTTACATGAAAAAGAAAAATTAATTAAGAAAGTAAAGTAAAGTAAAGTAAAGT 2379  
181 TTTAATTTTGAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 240  
2378 TTTAATTTTGAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2319  
241 TGTTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 300  
2318 TGTTTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2259  
301 TTTAATAAATAAGCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 360  
2258 TTTAATAAATAAGCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2199  
361 AAATACGTTTCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 420  
2198 AAATACGTTTCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2139  
421 AAGAACTATTGAGAAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 480  
2138 AAGAACTATTGAGAAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2079  
481 GTTGCAATGTTGTTCTTACTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 540  
2078 GTTGCAATGTTGTTCTTACTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2019  
541 NGAAATAA-CCTTAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 598  
2018 AGAATAATTAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1959  
599 TTGATTTTCTGTTCTTAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 658  
1958 TTGATTTTCTGTTCTTAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1859  
659 CCTAATTAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 686  
1898 CATTAATTAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1872  
RESULT 5  
AL358233  
LOCUS 122176 bp DNA linear HTG 10-JUL-2001  
DEFINITION Homo sapiens chromosome 1 clone RP5-835A17, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 8 unordered pieces.  
ACCESSION AL358233  
VERSION AL358233.3 GI:9797852  
KEYWORDS HTG; HTGS PHASE1; HTGS\_CANCELLED.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS McIay,K.  
TITLE Direct Submission



TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Homo sapiens chromosome 1, clone RP11-310J14  
Unpublished  
2 (bases 1 to 175046)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,  
Boguslavsky,L., Bouckhagter,B., Brown,A., Burkett,G., Castle,A.,  
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Deaellano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hages,B., Heathford,A., Horton,L.,  
Howland,U.C., Johnson,R., Jones,C., Kam,L., Karatas,A., Klein,J.,  
Landers,T., Lechoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,  
McPherson,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rognov,P., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testfay,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 175046)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Bouckhagter,B., Brown,A., Burkett,G.,  
Camporiano,A., Castle,A., Chapel,Y., Colangelo,M., Collins,S.,  
Collymore,A., Cooke,P., Deaellano,K., Dewar,K., Diaz,J.S.,  
Dodg,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hages,B., Heathford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karatas,A.,  
Klein,J., Lacroque,K., Lamazeres,R., Landers,T., Lechoczy,J.,  
Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPherson,R.,  
Meldrum,J., Menus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rognov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testfay,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2000 this sequence version replaced gi:6980310.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Genome Center

Center code: WIBR  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information

Center project name: U5491  
Center clone name: 310\_J\_14

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phred; version 0.960731

Consensus quality: 156135 bases at least Q40  
Consensus quality: 156373 bases at least Q30  
Consensus quality: 168980 bases at least Q20

Insert size: 185000; agarose-fp  
Insert size: 171746; sum-of-contigs

Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently

\* consists of 34 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 56: contig of 56 bp in length  
57 156: gap of 100 bp  
157 1191: contig of 1035 bp in length  
1191 1291: gap of 100 bp  
1291 1921: contig of 1141 bp in length  
1921 2432: gap of 100 bp  
2432 2532: gap of 100 bp  
2532 3742: contig of 1210 bp in length  
3742 3842: gap of 100 bp  
3842 4921: contig of 1079 bp in length  
4921 5021: gap of 100 bp  
5021 6073: contig of 1052 bp in length  
6073 6173: gap of 100 bp  
6173 7552: contig of 1379 bp in length  
7552 7652: gap of 100 bp  
7652 9001: contig of 1349 bp in length  
9001 9101: gap of 100 bp  
9101 10622: contig of 1521 bp in length  
10622 10723: gap of 100 bp  
10723 11835: contig of 1114 bp in length  
11835 11937: gap of 100 bp  
11937 13398: contig of 1462 bp in length  
13398 13498: gap of 100 bp  
13498 14949: contig of 1451 bp in length  
14949 15049: gap of 100 bp  
15049 15260: contig of 1211 bp in length  
15260 16360: gap of 100 bp  
16360 18261: contig of 1905 bp in length  
18261 18365: gap of 100 bp  
18365 19916: contig of 1551 bp in length  
19916 20016: gap of 100 bp  
20016 22939: contig of 2923 bp in length  
22939 23039: gap of 100 bp  
23039 26445: contig of 3406 bp in length  
26445 26545: gap of 100 bp  
26545 29327: contig of 2782 bp in length  
29327 29427: gap of 100 bp  
29427 32200: contig of 2773 bp in length  
32200 32300: gap of 100 bp  
32300 36738: contig of 4438 bp in length  
36738 36838: gap of 100 bp  
36838 43447: contig of 6609 bp in length  
43447 43547: gap of 100 bp  
43547 48525: contig of 4978 bp in length  
48525 48625: gap of 100 bp  
48625 52791: contig of 4166 bp in length  
52791 52891: gap of 100 bp  
52891 58391: contig of 5500 bp in length  
58391 58491: gap of 100 bp  
58491 64869: contig of 6378 bp in length  
64869 64969: gap of 100 bp  
64969 73369: contig of 8400 bp in length  
73369 73469: gap of 100 bp  
73469 81101: contig of 7632 bp in length  
81101 81201: gap of 100 bp  
81201 87754: contig of 6553 bp in length  
87754 87854: gap of 100 bp  
87854 95534: contig of 7680 bp in length  
95534 95634: gap of 100 bp  
95634 102797: contig of 7163 bp in length  
102797 102897: gap of 100 bp  
102897 111817: contig of 8920 bp in length  
111817 111917: gap of 100 bp  
111917 121160: contig of 9243 bp in length  
121160 121260: gap of 100 bp  
121260 140537: contig of 19277 bp in length  
140537 140637: gap of 100 bp



Db	Accession	Source	Organism	Reference Authors Title Journal	Comment
Db	42984	AAATACGTTTCGATAGCATGAAATGCAAAATTTTAAATTTTAAATCTACTAATTTT			
Qy	421	AAGAATCTATTGAGAAATTTGATTATGACATGAAATGACACATTAATTTACTGGCCAGCT			
Db	42924	AAGAACTATTGAGAAATTTGATTATGACATGAAATGACACATTAATTTACTGGCCAGCT			
Qy	481	GTTGGCAATGTTGTTTCTTACTAGTCTCCCAAGGAAATCTGTAACTGAATCTTCACG			
Db	42864	GTTGGCAATGTTGTTTCTTACTAGTCTCCCAAGGAAATCTGTAACTGAATCTTCACG			
Qy	541	NGAATAA-CTTAAATATATCTTTGTAGCCAAACAAA-CTTTTGTGTACATAGTTCT			
Db	42804	AGAAATATCTTTAAATATATCTTTGTAGCCAAACAAA-CTTTTGTGTACATAGTTCT			
Qy	599	TTGGAATTTTACTGTTCTCTTAATTTTATTTCTGAAATCTCAATTTTCCCAAGCATTAATAC			
Db	42744	TTGGAATTTTACTGTTCTCTTAATTTTATTTCTGAAATCTCAATTTTCCCAAGCATTAATAC			
Qy	659	CCATTTTAATCTTTGTATGACACAGTTGT			
Db	42684	CATA-TTAACTTTGTATGACACAGTTGT			
RESULT 7					
AL445183/c					
LOCUS	AL445183	193774 bp	DNA	linear	PRI 05-APR-2002
DEFINITION	Human DNA sequence from clone RP11-334A14 on chromosome 1, complete				
ACCESSION	AL445183				
VERSION	AL445183.19	GI:20068427			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Wallis, T				
TITLE	Direct Submission				
JOURNAL	Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humbrey@sanger.ac.uk				
COMMENT	On Apr 7, 2002 this sequence version replaced gi:17939714. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWSRPT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1 RP11-334A14 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBAC3.6.				
FEATURES					
source	location/Qualifiers				
	1..193774				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				

/chromosome="1"  
/clone="RP11-334A14"  
/clone.lib="RBC1-11.2"  
BASE COUNT 50815 a 45351 c 44031 g 53577 t  
ORIGIN

Query Match 92.7%; Score 638.4; DB 9; Length 193774;  
Best Local Similarity 98.1%; Pred. No. 9,1e-86;  
Matches 675; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

QY 1 ATTTAATTTCTAATTAACATTTCTTCAAGCATTAATTTATCTATATCTCACTGAA 60  
DB 98211 ATTTAATTTCTAATTAACATTTCTTCAAGCATTAATTTATCTATATCTCACTGAA 98152  
QY 61 TTTTAANAATAACCTTGTATTTGAAAAAATAGTAAATGAGATATTTAA 120  
DB 98151 TTTTAANAATAACCTTGTATTTGAAAAAATAGTAAATGAGATATTTAA 98092  
QY 121 ACTTACATGAAAAAGAAATTTATACAAAGAGCTGAGAACGTTAATTAATGAATGAGA 180  
DB 98091 ACTTACATGAAAAAGAAATTTATACAAAGAGCTGAGAACGTTAATTAATGAATGAGA 98032  
QY 181 TTATTAATTTGAAAACTGATCTGAAAAAGAACTTTATTTGTTCAATTAATTTATGATG 240  
DB 98031 TTATTAATTTGAAAACTGATCTGAAAAAGAACTTTATTTGTTCAATTAATTTATGATG 97972  
QY 241 TGTTTATGACTAATTAACGATTTTTCATTAATGAGAAACCATGTTAAATATTTTAT 300  
DB 97971 TGTTTATGACTAATTAACGATTTTTCATTAATGAGAAACCATGTTAAATATTTTAT 97912  
QY 301 TTTAAAAATAAGCCTGTGTTCAAGCTGTGATCATATTTCTTTTATTTGATTTGGGAAGA 360  
DB 97911 TTTAAAAATAAGCCTGTGTTCAAGCTGTGATCATATTTCTTTTATTTGATTTGGGAAGA 97852  
QY 361 AAATACTGTTTCTGATGACATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420  
DB 97851 AAATACTGTTTCTGATGACATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 97792  
QY 421 AAGAATCTATGAAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
DB 97791 AAGAATCTATGAAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 97732  
QY 481 GTTGGCATGTGTTCTTCTTACTTAACTGCTCCCAAGAAATCTTAACTGAATCTTGAGC 540  
DB 97731 GTTGGCATGTGTTCTTCTTACTTAACTGCTCCCAAGAAATCTTAACTGAATCTTGAGC 97672  
QY 541 NGAATTA-CCTTAATATATCTTTGTTAGCAAAACAA-CCTTTTGTGTTACATAGTTCT 598  
DB 97671 AGAATTAATCTTAAATATATCTTTGTTAGCAAAACAA-CCTTTTGTGTTACATAGTTCT 97612  
QY 599 TTGGATTTTACTGTTCCCTAATTTTATCTGAAATCTCCATTTTCCCAAGCATTAATTTAC 658  
DB 97611 TTGGATTTTACTGTTCCCTAATTTTATCTGAAATCTCCATTTTCCCAAGCATTAATTTAC 97552  
QY 659 CCTATTTAATCTTTGTTAGCAGATGCT 686  
DB 97551 CATATTTAATCTTTGTTAGCAGATGCT 97525

RESULT 8  
BC005911/c 1439 bp mRNA linear PRI 12-JUL-2001  
LOCUS Homo sapiens, steroid carrier protein 2, clone MGC:14505  
DEFINITION IMAGE:4287946, mRNA, complete cde.  
ACCESSION BC005911  
VERSION BC005911.1 GI:13543502  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1439)  
AUTHORS Strausberg,R.

## TITLE

Direct Submission  
Submitted (02-APR-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

## REMARK

## COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNU)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNU at: <http://image.llnl.gov>  
Series: IRAL Plate: 21 Row: n Column: 10  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 432978.  
Location/Qualifiers  
1..1439

## FEATURES

## source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LocusID:6342"  
/db\_xref="taxon:9606"  
/clone="MGC:14505 IMAGE:4287946"  
/tissue\_type="Brain, primitive neuroectodermal"  
/clone\_lib="NIH\_MGC\_56"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
74..505  
/codon\_start=1  
/product="steroid carrier protein 2"  
/protein\_id="AAH05911.1"  
/db\_xref="GI:13543503"  
/translation="MGPEPAASSFRHQIEAVPTSSADGFKANLVFEIEKLEERG  
EQGVKKIGGIFAEKVDGPGGKEATWVDVKNKGSLVPSDKKADCTTMAADSFYA  
LMTGKNKPOSAFPOGKLKITGNMGLMKLQNLQGNATL"

## CDS

## BASE COUNT

490 a 219 c 272 g 458 t

## ORIGIN

Query Match 92.7%; Score 636.8; DB 9; Length 1439;  
Best Local Similarity 98.0%; Pred. No. 7.2e-85;  
Matches 674; Conservative 0; Mismatches 11; Indels 3; Gaps 3;

QY 1 ATTTAATTTCTAATTAACATTTCTTCAAGCATTAATTTATCTATATCTCACTGAA 60  
DB 1398 ATTTAATTTCTAATTAACATTTCTTCAAGCATTAATTTATCTATATCTCACTGAA 1339  
QY 61 TTTTAANAATAACCTTGTATTTGAAAAAATAGTAAATGAGATATTTAA 120  
DB 1338 TTTTAANAATAACCTTGTATTTGAAAAAATAGTAAATGAGATATTTAA 1279  
QY 121 ACTTACATGAAAAAGAAATTTATACAAAGAGCTGAGAACGTTAATTAATGAATGAGA 180  
DB 1278 ACTTACATGAAAAAGAAATTTATACAAAGAGCTGAGAACGTTAATTAATGAATGAGA 1219  
QY 181 TTATTAATTTGAAAACTGATCTGAAAAAGAACTTTATTTGTTCAATTAATTTATGATG 240  
DB 1218 TTATTAATTTGAAAACTGATCTGAAAAAGAACTTTATTTGTTCAATTAATTTATGATG 1159  
QY 241 TGTTTATGACTAATTAACGATTTTTCATTAATGAGAAACCATGTTAAATATTTTAT 300  
DB 1158 TGTTTATGACTAATTAACGATTTTTCATTAATGAGAAACCATGTTAAATATTTTAT 1099  
QY 301 TTTAAAAATAAGCCTGTGTTCAAGCTGTGATCATATTTCTTTTATTTGATTTGGGAAGA 360  
DB 1098 TTTAAAAATAAGCCTGTGTTCAAGCTGTGATCATATTTCTTTTATTTGATTTGGGAAGA 1039

QY 361 AATAGTGTTCGTATGATGATGAATGCAAAATTTTATGATTTTAACTCNCATATTTT 420  
DB 1038 AATAGTGTTCGTATGATGATGAATGCAAAATTTTATGATTTTAACTCNCATATTTT 979  
QY 421 AAGAACTATTGAGAAATTTGATTAATGACATGAGTGCACAACTAATTTACTGCCAGCT 480  
DB 978 AAGAACTATTGAGAAATTTGATTAATGACATGAGTGCACAACTAATTTACTGCCAGCT 919  
QY 481 GTTGGCATTTGTGTTCTTACTTACTTGTCTCCCAAGAAAACTTTAACTGAATTTTCAAG 540  
DB 918 GTTGGCATTTGTGTTCTTACTTACTTGTCTCCCAAGAAAACTTTAACTGAATTTTCAAG 859  
QY 541 NGATAA-CCTTAATATATCTTTGTTAGCCAAACAAA-CTTTTGTGTTACATAGTTCT 598  
DB 858 AGAATAATCTTAATATATCTTTGTTAGCCAAACAAA-CTTTTGTGTTACATAGTTCT 799  
QY 599 TTGGATTTTACTGTTCTTAATTTTATTTCTGAAGTCCATTTTCCAGACCATTAATTTAC 658  
DB 798 TTGGATTTTACTGTTCTTAATTTTATTTCTGAAGTCCATTTTACCCAGACCATTAATTTAC 739  
QY 659 CCTATTATTAATTTGTTATGACAGTTGT 686  
DB 738 CATATTTAATCTTTGTTATGACAGTTGT 712

RESULT 9  
ARI46579 714 bp DNA linear PAT 08-AUG-2001  
LOCUS Sequence 19 from patent US 6218521.  
DEFINITION ARI46579  
ACCESSION ARI46579  
VERSION ARI46579.1 GI:15109768  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 714)  
AUTHORS Obata, Y.  
TITLE Isolated nucleic acid molecules associated with gastric cancer and methods for diagnosing and creating gastric cancer  
JOURNAL Patent: US 6218521-A 19 17-Apr-2001;  
FEATURES Location/Qualifiers  
source 1..714  
BASE COUNT 258 a 100 c 92 g 260 t 4 others  
ORIGIN

Query Match 89.1%; Score 611.8; DB 6; Length 714;  
Best Local Similarity 97.1%; Pred. No. 4,3e-81;  
Matches 671; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

QY 1 AATTTAATTTTCTATTAACAATCTTCTCAAGCATTTATTTATCTATATCTCACTGAA 60  
DB 15 AATTTAATTTTCTATTAACAATCTTCTCAAGCATTTATTTATCTATATCTCACTGAA 74  
QY 61 TTTTAAATAAATAAATCTTATTAAGAAAACTAGAAAAAGATTAAGAGATTAATTA 120  
DB 75 TTTTAAATAAATAAATCTTATTAAGAAAACTAGAAAAAGATTAAGAGATTAATTA 134  
QY 121 ACTTACATGAAGAAAGAAATTTATTAACAAGAGTGAAGAGTATTAATTAATGAATGAGA 180  
DB 135 ACTTACATGAAGAAAGAAATTTATTAACAAGAGTGAAGAGTATTAATTAATGAATGAGA 194  
QY 181 TTAATTAATTTGAAGAACTGATCTGAAGAAAGAACTTTATTTGTTCAATTAATTTATGATG 240  
DB 195 TTAATTAATTTGAAGAACTGATCTGAAGAAAGAACTTTATTTGTTCAATTAATTTATGATG 254  
QY 241 TGTGTTTATGATTAATTAATTAATTTTCAATTAAGAAAGAACTGTTAAATTAATTTTAT 300  
DB 255 TGTGTTTATGATTAATTAATTAATTTTCAATTAAGAAAGAAAGAACTGTTAAATTAATTTTAT 314  
QY 301 TTTAAATAAAGAGCTGTGTTCAAGCTGATCATATTTCTTTTATTTGATTTGGAGAGA 360  
DB 315 TTTAAATAAAGAGCTGTGTTCAAGCTGATCATATTTCTTTTATTTGATTTGGAGAGA 374

QY 361 AATAGTGTTCGTATGATGATGAATGCAAAATTTTATGATTTTAACTCNCATATTTT 420  
DB 375 AATAGTGTTCGTATGATGATGAATGCAAAATTTTATGATTTTAACTCNCATATTTT 434  
QY 421 AAGAACTATTGAGAAATTTGATTAATGACATGAGTGCACAACTAATTTACTGCCAGCT 480  
DB 435 AAGAACTATTGAGAAATTTGATTAATGACATGAGTGCACAACTAATTTACTGCCAGCT 494  
QY 481 GTTGGCATTTGTGTTCTTACTTACTTGTCTCCCAAGAAAACTTTAACTGAATTTTCAAG 539  
DB 495 GTTGGCATTTGTGTTCTTACTTACTTGTCTCCCAAGAAAACTTTAACTGAATTTTCAAG 554  
QY 540 CNGATAA-CCTTAATATATCTTTGTTAGCCAAACAAA-CTTTTGTGTTACATAGTTCT 597  
DB 555 CAGATAATCTTAATATATCTTTGTTAGCCAAACAAA-CTTTTGTGTTACATAGTTCT 614  
QY 598 TTT-GGATTTTACTGTTCTTAATTTTATTTCTGAAGTCCATTTTCCAGACCATTAATTT 656  
DB 615 TTGGGATTTTACTGTTCTTAATTTTATTTCTGAAGTCCATTTTACCCAGACCATTAATTT 674  
QY 657 ACCCTATTATTAATTTGTTATGACAGTTGT 687  
DB 675 ACCATA-TTAATCTTTGTTTGTGACAGTTGT 704

RESULT 10  
BD079828 714 bp DNA linear PAT 27-AUG-2002  
LOCUS BD079828  
DEFINITION Cancer-associated nucleic acids and polypeptides.  
ACCESSION BD079828  
VERSION BD079828.1 GI:22625431  
KEYWORDS UP 2001516009-A/494.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 714)  
AUTHORS Old, L.J., Scanlan, M.J., Stockert, E., Gure, A., Chen, Y.T., Gout, I., Ognate, M., Obata, Y., Freundschuh, M., Tureci, O. and Sahin, U.  
TITLE Cancer-associated nucleic acids and polypeptides  
JOURNAL Patent: JP 2001516009-A 494 25-SEP-2001;  
FEATURES Location/Qualifiers  
source 1..714  
BASE COUNT 258 a 100 c 92 g 260 t 4 others  
ORIGIN

Query Match 89.1%; Score 611.8; DB 6; Length 714;

QY 1 AATTTAATTTTCTATTAACAATCTTCTCAAGCATTTATTTATCTATATCTCACTGAA 60  
DB 15 AATTTAATTTTCTATTAACAATCTTCTCAAGCATTTATTTATCTATATCTCACTGAA 74  
QY 61 TTTTAAATAAATAAATCTTATTAAGAAAACTAGAAAAAGATTAAGAGATTAATTA 120  
DB 75 TTTTAAATAAATAAATCTTATTAAGAAAACTAGAAAAAGATTAAGAGATTAATTA 134  
QY 121 ACTTACATGAAGAAAGAAATTTATTAACAAGAGTGAAGAGTATTAATTAATGAATGAGA 180  
DB 135 ACTTACATGAAGAAAGAAATTTATTAACAAGAGTGAAGAGTATTAATTAATGAATGAGA 194  
QY 181 TTAATTAATTTGAAGAACTGATCTGAAGAAAGAACTTTATTTGTTCAATTAATTTATGATG 240  
DB 195 TTAATTAATTTGAAGAACTGATCTGAAGAAAGAACTTTATTTGTTCAATTAATTTATGATG 254  
QY 241 TGTGTTTATGATTAATTAATTAATTTTCAATTAAGAAAGAACTGTTAAATTAATTTTAT 300  
DB 255 TGTGTTTATGATTAATTAATTAATTTTCAATTAAGAAAGAAAGAACTGTTAAATTAATTTTAT 314  
QY 301 TTTAAATAAAGAGCTGTGTTCAAGCTGATCATATTTCTTTTATTTGATTTGGAGAGA 360  
DB 315 TTTAAATAAAGAGCTGTGTTCAAGCTGATCATATTTCTTTTATTTGATTTGGAGAGA 374



```

REFERENCE      1 (bases 1 to 1219)
AUTHORS        Yamamoto, R., Kallen, C.B., Babalola, G.O., Rennert, H.,
TITLE          Billheimer, J.T. and Straus, J.F. III.
                Cloning and expression of a cDNA encoding human sterol carrier
                protein 2
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 88 (2), 463-467 (1991)
MEDLINE        9110550
PUBMED        1703300
REFERENCE      2 (bases 1 to 1219)
AUTHORS        Vesa, J., Hellsten, E., Branoski, B.L., Emanuel, B.S., Billheimer, J.T.,
TITLE          Mead, S., Cowell, J.K., Straus, J.F. III, and Peltonen, L.
                Assignment of sterol carrier protein X/sterol carrier protein 2 to
                1p32 and exclusion as the causative gene for infantile neuronal
                ceroid lipofuscinosis
JOURNAL        Unpublished
COMMENT        On Dec 6, 1993 this sequence version replaced gi:432972.
FEATURES       location/Qualifiers
SOURCE         1..1219
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /map="1p32"
                /cissue_type="liver"
                1..1219
                /gene="SCP-2"
                /gene="SCP-2"
                /codon_start=1
                /product="sterol carrier protein-2"
                /protein_id="AA0559.1"
                /db_xref="GI:432973"
                /translation="MGPEPAASFFRTHQIEAVPTSSADGKALVKEIKKEEBEG
                EQPVKIGIGIPAKVDGPGGKEATWVDVNGGSVLPNSDKADCTTWDSDFLA
                LMTGKMPQSAFPOGKLKITGNMGLAMKLQVLQPGNAKL"
                84..143
                /gene="SCP-2"
                /gene="SCP-2"
                /product="unnamed"
                782..787
                /gene="SCP-2"
                /evidence=not_experimental
                1062..1067
                /gene="SCP-2"
                /evidence=not_experimental
                1107..1112
                /gene="SCP-2"
                /evidence=not_experimental
                1190..1195
                /gene="SCP-2"
                /evidence=not_experimental
                1219
                /evidence=experimental
                /evidence="SCP-2"
                /evidence=experimental
                195 c 255 g 367 t
BASE COUNT    402 a
ORIGIN
Query Match   64.1%; Score 440.4; DB 9; Length 1219;
Best Local Similarity 97.6%; Pred. No. 6.5e-56;
Matches 488; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

```

```

Oy 369 TTTCTGATAGCAAGAAATGCAAAATTTTATGATTTTATCTGCTAATTTTAAAGACTA 428
Db 1039 TTTCTGATAGCAAGAAATGCAAAATTTTATGATTTTATCTGCTAATTTTAAAGACTA 981
Oy 429 TTGAGAATTTGATTAATGACATGAGTGCACAACTAATTTACTGGCAGCTGTGGCAT 488
Db 980 TTGAGAATTTGATTAATGACATGAGTGCACAACTAATTTACTGGCAGCTGTGGCAT 921
Oy 489 TGTGTTCTTACTTATGTTCTCCCAAGAAACTCTTAACTGAATCTTTCAGCNGATTA- 547
Db 920 TGTGTTCTTACTTATGTTCTCCCAAGAAACTCTTAACTGAATCTTTCAGCNGATTA- 861
Oy 548 CCTAAATATACCTTGTATACCAACAAACAA-CTTTTGTTTATCATAGTCTTTGGAATT 606
Db 860 CCTAAATATACCTTGTATACCAACAAACAAAGCTTTTGTATCATAGTCTTTGGAATT 801
Oy 607 TACTGTTCTTAATTTTATTCGAAACCTCCATTTTCCCGACCATATTAACCTAATTTA 666
Db 800 TACTGTTCTTAATTTTATTCGAAACCTCCATTTTCCCGACCATATTAACCTAATTTA 742
Oy 667 ACTTTGTTATGACAGTTGT 686
Db 741 ACTTTGTTATGACAGTTGT 722

RESULT 13
LOCUS          ARI46582 843 bp DNA linear PAT 08-AUG-2001
DEFINITION    Sequence 22 from patent US 6218521.
ACCESSION     ARI46582
VERSION       ARI46582.1 GI:15109771
KEYWORDS      Unknown.
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 843)
AUTHORS        Obata, Y.
TITLE          Isolated nucleic acid molecules associated with gastric cancer and
                methods for diagnosing and treating gastric cancer
JOURNAL        Patent: US 6218521-A 22 17-APR-2001;
FEATURES       location/Qualifiers
SOURCE         1..843
                /organism="Unknown"
BASE COUNT    237 a 184 c 89 g 215 t 118 others
ORIGIN
Query Match   59.9%; Score 411.8; DB 6; Length 843;
Best Local Similarity 71.8%; Pred. No. 1.2e-51;
Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;

```

Db 315 TTTAAATTAACCCGNTTCCAAACCCNGATCANAATTCCTTNAATTTGGATTGGGAAA 374  
 Qy 361 AAATCTGTTCTGATAGATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420  
 Db 375 AAATNCNGTCCNNATACCCNNAANNNGCAAAATTTTAAATTTTAAACCCCTANTTTT 434  
 Qy 421 AAGACTATTGAGAAATGATTAATGACATGAGTGAACACTAATTAATGAGGAGCT 480  
 Db 435 AAAAATCTATNGAAANTNGATTANNAGCTGAATTCG-CAACCTANTTNCNGGACACN 493  
 Qy 481 GTTGCAATTTGTTGTTTCTTACTAGTTCTCCAGGAAAATCTTAACTGAATCTTCAGC 540  
 Db 494 GTGGCCTNTGNTTCTTACTTANTCCCCCAAGGAAANNCTTAANGAANGCTCCNC 553  
 Qy 541 NGAATTAACCTTAATATATCTTGTAGCCAAAC-AAAATCTTTTGTATGATGTTCT 598  
 Db 554 AAAATTAACCTTAATATATCTTGTAGCCAAACAAACCTTTTNGTTTACNTANTCT 613  
 Qy 599 TTGCAATTTACTGTTCTTAAATTTTATCTGAACCTCAATTTTCCGAGACCATTAATTC 658  
 Db 614 TGGGATTTAACGGGTCCCAATTTTATCNGAACCAATTTTCCCAACCATANTTAC 673  
 Qy 659 CCAATTTAACTTTGTTTATGACAGTTGTT 687  
 Db 674 CAT-TTTACTTGTGAAGCNCAGTNGTT 701

## RESULT 14

BD079831 843 bp DNA linear PAT 27-AUG-2002  
 LOCUS Cancer-associated nucleic acids and polypeptides.  
 DEFINITION BD079831  
 ACCESSION BD079831 GI:22625434  
 VERSION JP 2001516009-A/497.  
 KEYWORDS Homo sapiens (human).  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 843)  
 AUTHORS Old,L.J., Scanlan,M.J., Stockert,E., Gure,A., Chen,Y.T., Gout,I.,  
 Ognare,M., Odata,Y., Pfreundschuh,M., Tureci,O. and Sahin,U.  
 Cancer-associated nucleic acids and polypeptides  
 Patent: JP 2001516009-A 497 25-SEP-2001;  
 LUDWIG INSTITUTE FOR CANCER RESEARCH  
 COMMENT OS Homo sapiens (human)  
 PN JP 2001516009-A/497  
 PD 25-SEP-2001  
 PR 15-JUL-1998 JP 2000503425  
 PR 17-JUL-1997 US 08/896164,10-OCT-1997 US 60/061599 PR  
 10-OCT-1997 US 60/061765,10-OCT-1997 US 08/948705 PR  
 11-OCT-1997 GB 9721697.2,22-JUN-1998 US 09/102322 PI LLOYD  
 J OLD,MATTHEW J SCANLAN,ELISABETH STOCKERT,ALI GURE,YAO PI TSENG  
 CHEN,  
 PI IVAN GOUT,MICHAEL O'HARE,YUICHI OBATA,MICHAEL PFREUNDSCHEID, PI  
 OZLEM TURECI,  
 PI UGUR SAHIN

FEATURES  
 source Location/Qualifiers  
 1..843 /organism="Homo sapiens (human)".  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 237 a 184 c 89 g 215 t 118 others  
 ORIGIN

Query Match 59.9%; Score 411.8; DB 6; Length 843;  
 Best Local Similarity 71.8%; Pred. No. 1.2e-51;  
 Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;

Qy 1 ATTTAAATTTTCCTTAAACATCTCTGCAAGCATTTATATCTATATCTCATGAA 60  
 Db 15 ATTTAAATTTTCCTTAAATGATCTCTGCAAGCATTTATATCTATATCTCATGAA 74  
 Qy 61 TTTTAAATAATTAACATTAATGATTAAGAAAATCTAGAAAAAGATTAATGCAATTAAT 120  
 Db 75 TTTTAAATAATTAACATTTTNTTTTAAAAAACCCTGAAAAAATAATGCAATTAATTA 134  
 Qy 121 ACTTACATGAAAAAGAAAAATTAATACAAAGGAGTGAAGCTTAAATTAATGAATGGA 180  
 Db 135 CTTTCTGAAAAAGAAAAATTTTACAAAGGAGCAGAAACCTTAAATTAATGAAATTA 194  
 Qy 181 TTAATAATTTGAAAAATGCAATGCAAACTTAATTTGTTCAATTAATTTGTTGATG 240  
 Db 195 TTAATNTTNGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 254  
 Qy 241 TGTTTTATGACTAATTAACATGATTTTCAATTAAGAAAAACCATGTTAAAAATTTTAT 300  
 Db 255 NNTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 314  
 Qy 301 TTTTAAATAATTAACATGATTTTCAATTAATTAATTAATTAATTAATTAATTAAT 360  
 Db 315 TTTTAAATAATTAACATGATTTTCAATTAATTAATTAATTAATTAATTAATTAAT 374  
 Qy 361 AAATCTGTTCTGATAGATGAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420  
 Db 375 AAATNCNGTCCNNATACCCNNAANNNGCAAAATTTTAAATTTTAAACCCCTANTTTT 434  
 Qy 421 AAGACTATTGAGAAATGATTAATGACATGAGTGAACACTAATTAATGAGGAGCT 480  
 Db 435 AAAAATCTATNGAAANTNGATTANNAGCTGAATTCG-CAACCTANTTNCNGGACACN 493  
 Qy 481 GTTGCAATTTGTTGTTTCTTACTAGTTCTCCAGGAAAATCTTAACTGAATCTTCAGC 540  
 Db 494 GTGGCCTNTGNTTCTTACTTANTCCCCCAAGGAAANNCTTAANGAANGCTCCNC 553  
 Qy 541 NGAATTAACCTTAATATATCTTGTAGCCAAAC-AAAATCTTTTGTATGATGTTCT 598  
 Db 554 AAAATTAACCTTAATATATCTTGTAGCCAAACAAACCTTTTNGTTTACNTANTCT 613  
 Qy 599 TTGCAATTTACTGTTCTTAAATTTTATCTGAACCTCAATTTTCCGAGACCATTAATTC 658  
 Db 614 TGGGATTTAACGGGTCCCAATTTTATCNGAACCAATTTTCCCAACCATANTTAC 673  
 Qy 659 CCAATTTAACTTTGTTTATGACAGTTGTT 687  
 Db 674 CAT-TTTACTTGTGAAGCNCAGTNGTT 701

## RESULT 15

HUMSCP2B/c 1500 bp mRNA linear PRI 06-DEC-1993  
 LOCUS Human sterol carrier protein 2 mRNA, complete cds.  
 DEFINITION M75884  
 ACCESSION M75884.1 GI:432976  
 VERSION sterol carrier protein-2.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1500)  
 AUTHORS He,Z., Yamamoto,R., Furch,B.E., Schanz,L.J., Maylor,S.L.,  
 George,H., Billheimer,J.T. and Strause,J.P. III.  
 CDNA encoding members of a family of proteins related to human  
 sterol carrier protein 2 and assignment of the gene to human  
 chromosome 1 p21----pter  
 JOURNAL DNA Cell Biol. 10 (8), 559-569 (1991)  
 MEDLINE 92029618  
 PUBMED 1718316

REFERENCE 2 (bases 1 to 1500)  
 AUTHORS Vesa,J., Hellesten,E., Branoski,B.L., Emanuel,B.S., Billheimer,J.T.,  
 Mead,S., Cowell,J.K., Strassus,J.F.III, and Peltonen,L.  
 TITLE Assignment of sterol carrier protein X/sterol carrier protein 2 to  
 1p32 and exclusion as the causative gene for infantile neuronal  
 ceroid lipofuscinosis  
 JOURNAL Unpublished  
 COMMENT On Dec 6, 1993 this sequence version replaced gi:337996.  
 FEATURES  
 Location/Qualifiers  
 source  
 1..1500  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /map="1p32"  
 /issue\_type="1iver"  
 /clone\_id="Clontech HL-1001b;HL-1115b"  
 1..1500  
 /gene="SCP-2"  
 82..951  
 /gene="SCP-2"  
 /codon\_start=1  
 /product="sterol carrier protein-2"  
 /protein\_id="AA03558.1"  
 /db\_xref="GI:432977"

gene  
 CDS  
 1218..1223  
 /translation="NM0DLPSSFEKSIIMVGFPMSEKAAKCYEKSGLTPNDIVY  
 ELHDCSTNELITYEALGLCEPQGLVDRGDNVYGGKVINPISGLISKPHLGT  
 GLAQCAELCQLRGKAGKROVPGAKVALQHNLGIGAVVTLVYMGFPEAASSFRHQ  
 IEAVPTSSASDGFANLVFKEIEKLEBEGQFVKIGIIFAFVKDGPGEKATWV  
 DVKNGKSVLPNSDKKADCTITMADSDFLALMTGKNPQSAFQGLKLTGNMGLAK  
 LONLQPGNAKL"  
 1218..1223  
 /gene="SCP-2"  
 /note="no polyA or polyA signal was found at the end of  
 this cDNA clone, indicating that the mRNA sequence may be  
 partial at the 3' end"  
 /evidence=not experimental

BASE COUNT 482 a 254 c 342 g 422 t  
 ORIGIN

Query Match 43.1%; Score 296.4; DB 9; Length 1500;  
 Best Local Similarity 96.8%; Pred. No. 9.9e-35;  
 Matches 333; Conservative 0; Mismatches 8; Indels 3; Gaps 3;  
 QY 345 TTTTGATTTGGGAAGAAATCTGTTCTGTGATGATGAATGCAAAATTTTGATTTT 404  
 |||||  
 DB 1500 TTTTGATTTGGGAAGAAATCTGTTCTGTGATGATGAATGCAAAATTTTGATTTT 1441  
 QY 405 TAATCTGCTAATTTTAAGACTATTGGAATGATTAATGACATGAAAGTCACAC 464  
 |||||  
 DB 1440 TAATCTGCTAATTTTAAGACTATTGGAATGATTAATGACATGAAAGTCACAC 1381  
 QY 465 TAATTACTGGCAGCTGTGGCATGTGTTCTTACTAGTTCTCCAGGAAACTCTT 524  
 |||||  
 DB 1380 TAATTACTGGCAGCTGTGGCATGTGTTCTTACTAGTTCTCCAGGAAACTCTT 1321  
 QY 525 AAATGTAATCTTCAGCNGAATAA-CCTTAATAATATTGTTAGCCAAACAAA-CTTTT 582  
 |||||  
 DB 1320 AAATGTAATCTTCAGCNGAATAA-CCTTAATAATATTGTTAGCCAAACAAA-CTTTT 1261  
 QY 583 TTGTTTACATAGTCTTGTGATTTTACTGTTCTTAATTTTATCTGAAACTCCATTTTC 642  
 |||||  
 DB 1260 TTGTTTACATAGTCTTGTGATTTTACTGTTCTTAATTTTATCTGAAACTCCATTTTAC 1201  
 QY 643 CCCAGACCATTAATACCTTAATTTACTTTGTTATGACAGTTGT 686  
 |||||  
 DB 1200 CCCAGACCATTAATACCATTA-TTAACCTTGTATGACAGTTGT 1158

Search completed: November 27, 2003, 10:02:34  
 Job time : 2784.58 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:55:53 ; Search time 1851.59 Seconds  
(without alignments)  
9017.749 Million cell updates/sec

Title: US-09-835-992a-20

Perfect score: 687  
Sequence: 1 attctatcttctcattcaaac.....cttctgtatgcacagcttct 687

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estsba:\*  
2: em\_estsba:\*  
3: em\_estsba:\*  
4: em\_estsba:\*  
5: em\_estsba:\*  
6: em\_estsba:\*  
7: em\_estsba:\*  
8: em\_estsba:\*  
9: em\_estsba:\*  
10: em\_estsba:\*  
11: em\_estsba:\*  
12: em\_estsba:\*  
13: em\_estsba:\*  
14: em\_estsba:\*  
15: em\_estsba:\*  
16: em\_estsba:\*  
17: em\_estsba:\*  
18: em\_estsba:\*  
19: em\_estsba:\*  
20: em\_estsba:\*  
21: em\_estsba:\*  
22: em\_estsba:\*  
23: em\_estsba:\*  
24: em\_estsba:\*  
25: em\_estsba:\*  
26: em\_estsba:\*  
27: em\_estsba:\*  
28: em\_estsba:\*  
29: em\_estsba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	639	93.0	769	BQ014192 UI-H-ED1-
2	638.4	92.8	12	BM997078 UI-H-ED0-
3	637.6	92.8	837	A1566109 tns3e08.x
4	636.8	92.7	738	BQ045161 UI-CF-EN1

5	635.8	92.5	712	13	BU619082
6	635.8	92.5	751	12	BO008197
7	635.8	92.5	759	13	BU619112
8	635.4	92.5	714	9	A1675901
9	635.2	92.5	756	12	BM985376
10	635.2	92.5	762	13	BU627147
11	634.4	92.3	770	9	A1826287
12	632	92.0	721	12	BM968746
13	632	92.0	734	14	CA424156
14	632	92.0	786	9	AW052045
15	630.2	91.7	750	12	BO009270
16	629.8	91.7	734	13	BU933572
17	629.4	91.6	793	9	A1565988
18	624	90.8	933	9	AL578584
19	623	90.7	672	12	BO016778
20	623	90.7	905	13	BK414370
21	617.8	89.9	901	13	BK392565
22	616.6	89.8	822	9	AV729461
23	616	89.7	821	9	A1640146
24	609.4	88.7	876	9	A1831751
25	606	88.2	671	9	A1683310
26	605.4	88.1	784	12	B1861076
27	600.8	87.5	699	10	BE739622
28	599.2	87.2	824	10	BE738457
29	592	86.2	728	9	A1831053
30	591.4	86.1	1032	9	AL551280
31	590	85.9	840	9	AA659242
32	589.4	85.8	648	12	BM511898
33	587.2	85.5	641	9	AM173415
34	586.8	85.4	749	9	A1765940
35	583.4	84.9	616	9	AM611897
36	582.8	84.8	820	10	BP966918
37	582.4	84.8	660	9	A1751686
38	577.6	84.1	673	9	AM512844
39	570	83.0	702	9	A1431843
40	569.6	82.9	637	13	BQ549593
41	569	82.8	727	9	A1636002
42	568	82.7	744	10	BG569206
43	566.4	82.4	627	9	A1224925
44	565.4	82.3	753	10	BE748610
45	563.4	82.0	623	12	BM511625

## ALIGNMENTS

RESULT 1  
BQ014192  
LOCUS  
DEFINITION  
UI-H-ED1-ax8-g-24-0-UI-B1\_NCI\_CGAP\_ED1 Homo sapiens CDNA clone  
IMAGE:5833007 3', mRNA sequence.  
BQ014192  
BQ014192.1 GI:19739093  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 769)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS  
TITLE  
Tumor Gene Index  
JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Straube, Ph.D.  
Email: [cgapbs-romail.nih.gov](mailto:cgapbs-romail.nih.gov)  
Tissue Procurement: Dr. Jose Mercuende  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found  
through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>  
The following repetitive elements were found in this CDNA  
sequence: 317-343, >AT\_richlow\_complexity (matched complement)

Seq primer: M13 FORWARD  
POLYA=Yes

FEATURES  
Source

Location/Qualifiers  
1..769  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5833007"  
/tissue\_type="Chondrosarcoma"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP ED1"  
/note="Organ: Left Pubic Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP ED1 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma cell line C5. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is GCTCAAGGCT.  
TAG\_LIB=UI-H-ED1  
TAG\_TISSUE=chondrosarcoma  
TAG\_SEQ=CGTCAAGGCT"

BASE COUNT 272 a 110 c 100 g 285 t 2 others  
ORIGIN

Query Match 93.0%; Score 639; DB 12; Length 769;  
Best Local Similarity 98.1%; Pred. No. 3,7e-66;  
Matches 675; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

1 ATTTAATTTCTATTAACATTTCTCTCAAGCATTTTATCTATATCTCACTGAA 60  
|||||  
33 ATTTAATTTCTATTAACATTTCTCTCAAGCATTTTATCTATATCTCACTGAA 92  
|||||

61 TTTTAAATAATACATTAGATTGAAAAAAGTAAAGTAAATGACATTAATTA 120  
|||||  
93 TTTTAAATAATACATTAGATTGAAAAAAGTAAAGTAAATGACATTAATTA 152  
|||||

121 ACTTACTGAAAAAGAAATTTAACAAGAGATGAAAGCTATTAATTTGAATAGA 180  
|||||  
153 ACTTACTGAAAAAGAAATTTAACAAGAGATGAAAGCTATTAATTTGAATAGA 212  
|||||

181 TTAATATTTGAAAACTGATCTGAAAGCAAACTTATTTGTTCAATTAATCTAATGATGG 240  
|||||  
213 TTAATATTTGAAAACTGATCTGAAAGCAAACTTATTTGTTCAATTAATCTAATGATGG 272  
|||||

241 TGTATTATGACTAATACAGTATTTTCAATTAAGAAACCATGTTAAAAATATTTTAT 300  
|||||  
273 TGTATTATGACTAATACAGTATTTTCAATTAAGAAACCATGTTAAAAATATTTTAT 332  
|||||

301 TTTTAAATAATAGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTGATTTGGAGAGA 360  
|||||  
333 TTTTAAATAATAGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTGATTTGGAGAGA 392  
|||||

361 AAATAGCTGTTTCTGATGATGAAAGTAAATTTTGAATTTTATCTCNCATATTTT 420  
|||||  
393 AAATAGCTGTTTCTGATGATGAAAGTAAATTTTGAATTTTATCTCNCATATTTT 452  
|||||

421 AAGAACTATTGAGAAATTTGATTAATGACATGAAGTGAACAACATAATTAATGAGCCAGCT 480  
|||||  
453 AAGAACTATTGAGAAATTTGATTAATGACATGAAGTGAACAACATAATTAATGAGCCAGCT 512  
|||||

481 GTTGGCAATTGTCTTTCTTACTAGTTCTCCCAAGAAAACTTTAAACTGAATCTTTAGC 540  
|||||  
513 GTTGGCAATTGTCTTTCTTACTAGTTCTCCCAAGAAAACTTTAAACTGAATCTTTAGC 572  
|||||

541 AAGAAATAA-CCTTAAATATCTGTTGATGCAAAACAAA-CCTTTTGTGTTAATCATAGTTCT 598  
|||||

Db 573 AGAATATCTCTTAATATATCTTTGTAAGCAAAACAAAGCTTTTGTGTACTAGTCT 632  
|||||

Qy 599 TTGATTTTACTGTTCTCTAATTTTATTTGTAAGCAAACTCCATTTTCCCGACCATTAATTAC 658  
|||||

Db 633 TTGATTTTACTGTTCTCTAATTTTATTTGTAAGCAAACTCCATTTTACCCGACCATTAATTAC 692  
|||||

Qy 659 CCTATTAACTTTGTTATGACAGTTGT 686  
|||||

Db 693 CATATTTAATTTGTTATGACAGTTGT 719  
|||||

RESULT 2  
BM997078  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BM997078 761 bp mRNA linear EST 26-MAR-2002  
UI-H-ED0-axo-f-03-0-UI.s1 NCI CGAP ED0 Homo sapiens cDNA clone  
IMAGE:5831426 3', mRNA sequence.  
BM997078  
BM997078.1 GI:19721979  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 761)  
NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arived by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILN.ac: http://image.llnl.gov  
The following repetitive elements were found in this cDNA  
sequence: 316342>AT richlow\_complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
Source

Location/Qualifiers  
1..761  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5831426"  
/tissue\_type="Chondrosarcoma"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP ED0"  
/note="Organ: Left Pubic Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP ED0 is a cDNA library containing the following tissue(s): Chondrosarcoma cell line C5. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is GCTCAAGGCT.  
TAG\_LIB=UI-H-ED0  
TAG\_TISSUE=chondrosarcoma  
TAG\_SEQ=CGTCAAGGCT"

BASE COUNT 269 a 108 c 98 g 285 t 1 others  
ORIGIN

Query Match 92.9%; Score 638.4; DB 12; Length 761;  
Best Local Similarity 98.1%; Pred. No. 4.4e-66;

Matches 675; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

QY 1 ATTTAATTTTCTATTAACATTTCTTCAAGATATTTATCCATATCTCACAGAA 60  
 DB 32 ATTTAATTTTCTATTAACATTTCTTCAAGATATTTATCCATATCTCACAGAA 91  
 QY 61 TTTTAAATAATTAACATTAATAGAAAAGTAAATGAGATTAATTA 120  
 DB 92 TTTTAAATAATTAACATTAATAGAAAAGTAAATGAGATTAATTA 151  
 QY 121 ACTTACATGAAGAAAGAAATTAATACAAAGACTGAGAACTTATTAATGAATAGA 180  
 DB 152 ACTTACATGAAGAAAGAAATTAATACAAAGACTGAGAACTTATTAATGAATAGA 211  
 QY 181 TTTAATATTTAAATTCGATCTGAGAAACCTTATTTCTCAATTTCTTATATGAG 240  
 DB 212 TTTAATATTTAAATTCGATCTGAGAAACCTTATTTCTCAATTTCTTATATGAG 271  
 QY 241 TGTATTAATGACTATATACATGATTTTCAATTAAGAAACCATGTTAAATATTTTAT 300  
 DB 272 TGTATTAATGACTATATACATGATTTTCAATTAAGAAACCATGTTAAATATTTTAT 331  
 QY 301 TTTAATAATTAAGCTGTGTTCAAGCTGTGATCATATTTCTTTAATTTGATTTGGAGA 360  
 DB 332 TTTAATAATTAAGCTGTGTTCAAGCTGTGATCATATTTCTTTAATTTGATTTGGAGA 391  
 QY 361 AATATCTGTTTCTGATAGCATGAATGCAAAATTTTGAATTTTATCTCNCATATTTT 420  
 DB 392 AATATCTGTTTCTGATAGCATGAATGCAAAATTTTGAATTTTATCTCNCATATTTT 451  
 QY 421 AAGACATTTAGAAATTTGATTAATGACATGAGTGCACAACTAATTAATCTGCGCAGCT 480  
 DB 452 AAGACATTTAGAAATTTGATTAATGACATGAGTGCACAACTAATTAATCTGCGCAGCT 511  
 QY 481 GTTGGCATTTGTTCTTACTAGTTCTCCAGAGAAACCTTAACTGAATCTTTCAGC 540  
 DB 512 GTTGGCATTTGTTCTTACTAGTTCTCCAGAGAAACCTTAACTGAATCTTTCAGC 571  
 QY 541 NGATAA-CCTTAATATTAATTTGTTAGCCAAAGAAA-CTTTTGTGTTTACATAGTTCT 598  
 DB 572 AGAATATCTTAAATTAATTAATTTGTTAGCCAAAGAAA-CTTTTGTGTTTACATAGTTCT 631  
 QY 599 TTTGATTTTACTGTTCTTAAATTTTATCTGAAACTCCATTTTCCGAGCAATATTTAC 658  
 DB 632 TTTGATTTTACTGTTCTTAAATTTTATCTGAAACTCCATTTTCCGAGCAATATTTAC 691  
 QY 659 CCTAATTAATCTTGTATGACAGTTGT 686  
 DB 692 CATATTTAATCTTGTATGACAGTTGT 718

RESULT 3  
 A1566109 837 bp mRNA linear EST 12-MAY-1999  
 LOCUS t053408.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:212124 3'  
 DEFINITION similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR  
 (HUMAN); mRNA sequence.  
 ACCESSION A1566109.1 GI:4524561  
 VERSION A1566109  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 837)  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIN at:  
[www-bio.liml.gov/bxrp/image/image.html](http://www-bio.liml.gov/bxrp/image/image.html)  
 Insert Length: 619 Std Error: 0.00  
 Seq primer: -40UP from Glibco  
 High quality sequence stop: 465  
 FOLTA=No.

FEATURES  
 source Location/Qualifiers  
 1..837  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:212124"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney; Vector: pRTD-Pac (Pharmacia) with  
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
 Plasmid DNA from the normalized library NCI-CCAP Kid3 was  
 prepared, and 88 circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneids 1323376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Facina Bonaldo."

BASE COUNT 300 a 129 c 118 g 285 t 5 others  
 ORIGIN

Query Match 92.8%; Score 637.6; DB 9; Length 837;  
 Best Local Similarity 97.5%; Pred. No. 5.2e-66;  
 Matches 654; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 1 ATTTAATTTTCTATTAACATTTCTTCAAGATATTTATCCATATCTCACAGAA 60  
 DB 18 ATTTAATTTTCTATTAACATTTCTTCAAGATATTTATCCATATCTCACAGAA 77  
 QY 61 TTTTAAATAATTAACATTAATAGAAAAGTAAATGAGATTAATTA 120  
 DB 78 TTTTAAATAATTAACATTAATAGAAAAGTAAATGAGATTAATTA 137  
 QY 121 ACTTACATGAAGAAAGAAATTAATACAAAGACTGAGAACTTATTAATGAATAGA 180  
 DB 138 ACTTACATGAAGAAAGAAATTAATACAAAGACTGAGAACTTATTAATGAATAGA 197  
 QY 181 TTTAATATTTGAAATTCGATCTGAGAAACCTTATTTCTCAATTTCTTATATGATG 240  
 DB 198 TTTAATATTTGAAATTCGATCTGAGAAACCTTATTTCTCAATTTCTTATATGATG 257  
 QY 241 TGTATTAATGACTATATACATGATTTTCAATTAAGAAACCATGTTAAATATTTTAT 300  
 DB 258 TGTATTAATGACTATATACATGATTTTCAATTAAGAAACCATGTTAAATATTTTAT 317  
 QY 301 TTTAATAATTAAGCTGTGTTCAAGCTGTGATCATATTTCTTTAATTTGATTTGGAGA 360  
 DB 318 TTTAATAATTAAGCTGTGTTCAAGCTGTGATCATATTTCTTTAATTTGATTTGGAGA 377  
 QY 361 AATATCTGTTTCTGATAGCATGAATGCAAAATTTTGAATTTTATCTCNCATATTTT 420  
 DB 378 AATATCTGTTTCTGATAGCATGAATGCAAAATTTTGAATTTTATCTCNCATATTTT 437  
 QY 421 AAGACATTTAGAAATTTGATTAATGACATGAGTGCACAACTAATTAATCTGCGCAGCT 480  
 DB 438 AAGACATTTAGAAATTTGATTAATGACATGAGTGCACAACTAATTAATCTGCGCAGCT 497  
 QY 481 GTTGGCATTTGTTCTTACTAGTTCTCCAGAGAAACCTTAACTGAATCTTTCAGC 540  
 DB 498 GTTGGCATTTGTTCTTACTAGTTCTCCAGAGAAACCTTAACTGAATCTTTCAGC 557  
 QY 541 NGATAA-CCTTAATATTAATTTGATGCAAACTTTTGTGTTTACATAGTTCTT 599

Db 558 AGAATAATCCCTAATAATACCTTGGTAGCAAAAGAGCTNNNGTTATACATAGTTCCT 617  
 QY 600 TGGATTTTACTGTTCCCTAATTTATTTGGAACCTCCTTTTCCCGAGACCATATATACC 659  
 Db 618 TGGATTTTACTGTTCCCTAATTTATTTGGAACCTCCTTTTCCCGAGACCATATATACC 677  
 QY 660 CTAATTAAGT 670  
 Db 678 ATATTTACTT 688  
 RESULT 4  
 B0045161 738 bp mRNA linear EST 21-FEB-2003  
 LOCUS  
 DEFINITION UI-CF-EN1-a-17-0-UI.61 UI-CF-EN1 Homo sapiens cDNA clone  
 ACCESSION B0045161  
 VERSION B0045161.1 GI:19796248  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 738)  
 TITLE Ronaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT  
 Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccray@uiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com) or from Open Biosystems  
 (www.openbiosystems.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 316-342, >AT-rich#low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=yes  
 Location/Qualifiers  
 1. 738  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-EN1-a-17-0-UI"  
 /tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
 Cells"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-CF-EN1"  
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-EN1 is a normalized cDNA library containing the  
 following tissue(s): Primary Lung Cystic Fibrosis  
 Epithelial Cells. The library was constructed according to  
 Ronaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not  
 I, and cloned directionally into pT73-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is CTGCTCAGGT.

TAG LIB=UI-CF-EN1  
 TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
 6hr to LPS 24h  
 TAG\_SEQ=CTGCTCAGGT"  
 BASE COUNT 260 a 102 c 97 g 278 t 1 others  
 ORIGIN  
 Query Match 92.7%; Score 636.8; DB 12; Length 738;  
 Best Local Similarity 98.0%; Pred. No. 6.8e-66;  
 Matches 674; Conservative 0; Mismatches 11; Indels 3; Gaps 3;  
 QY 1 ATTTTATTTTCTATTAAACATCTCTCAAGACATTTATTTTATCTCACTGAA 60  
 Db 32 ATTTTATTTTCTATTAAACATCTCTCAAGACATTTATTTTATCTCACTGAA 91  
 QY 61 TTTTAAATAATACATTAGTATTGAAAACTGAAAAAGATNAATGACATTAATTA 120  
 Db 92 TTTTAAATAATACATTAGTATTGAAAACTGAAAAAGATNAATGACATTAATTA 151  
 QY 121 ACTTACATGAAAAAGAAAAATTTATACAAAGACAGAACTTAAATGAAAGAGA 180  
 Db 152 ACTTACATGAAAAAGAAAAATTTATACAAAGACAGAACTTAAATGAAAGAGA 211  
 QY 181 TTATTAATTTGAAAACTGCACTGAAAAAGAACTTTATGTTCAATTATTTCTAATGATG 240  
 Db 212 TTATTAATTTGAAAACTGCACTGAAAAAGAACTTTATGTTCAATTATTTCTAATGATG 271  
 QY 241 TGTTTTATGACATAACACTGATTTTCAATTAAGAAACCCATGTTAAAAATTTTAT 300  
 Db 272 TGTTTTATGACATAACACTGATTTTCAATTAAGAAACCCATGTTAAAAATTTTAT 331  
 QY 301 TTTTAAATAATGACCTGCTGCTCAAGCTGATGATATTTCTTTATTTGATTTGGGAAG 360  
 Db 332 TTTTAAATAATGACCTGCTGCTCAAGCTGATGATATTTCTTTATTTGATTTGGGAAG 391  
 QY 361 AATATCTGTTTCGATAGACATGAATGCAAAATTTTGAATTTTATCTCCTAATTTT 420  
 Db 392 AATATCTGTTTCGATAGACATGAATGCAAAATTTTGAATTTTATCTCCTAATTTT 451  
 QY 421 AAGAACTATTGAAAAATGATTAATGACATGAAGTGCACAACTAATTAATGCGCAGCT 480  
 Db 452 AAGAACTATTGAAAAATGATTAATGACATGAAGTGCACAACTAATTAATGCGCAGCT 511  
 QY 481 GTTGGCATGCTGTTTCTTCTAGTCTCCCAAGAAACCTCTTAACGAACTCTCAG 540  
 Db 512 GTTGGCATGCTGTTTCTTCTAGTCTCCCAAGAAACCTCTTAACGAACTCTCAG 571  
 QY 541 NGATAAT-A-CCTTAATATATCTTTGTTAGCAAAACAAA-CCTTTTGTGTTACATAGTCT 598  
 Db 572 AGAATAATCTTAAATATATCTTTGTTAGCAAAACAAA-CCTTTTGTGTTACATAGTCT 631  
 QY 599 TTTGATTTTACTGTTCTTAATTTTATTTGAAAACTCAATTTTCCCGACCATTAATAC 658  
 Db 632 TTTGATTTTACTGTTCTTAATTTTATTTGAAAACTCAATTTTCCCGACCATTAATAC 691  
 QY 659 CCAATTAAGTCTTTGTTATGACAGTGT 686  
 Db 692 CATA-TTAACTTTGTTATGACAGTGT 718  
 RESULT 5  
 B0619082 712 bp mRNA linear EST 23-SEP-2002  
 LOCUS  
 DEFINITION UI-H-FH1-bfm-f-06-0-UI.61 NC1 CGAP FH1 Homo sapiens cDNA clone  
 ACCESSION B0619082  
 VERSION B0619082.1 GI:23285297  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 712)



line C85. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAAGGCT.

TAG LIB=UI-H-ED1  
TAG TISSUE=chondrosarcoma  
TAG SEQ=CGTCAAGGCT"

BASE COUNT 266 a 104 c 98 g 282 t 1 others  
ORIGIN

Query Match 92.5%; Score 635.8; DB 12; Length 751;  
Best Local Similarity 97.8%; Pred. No. 8.9e-66;  
Matches 674; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

```

QY 1 ATTTAATTTTCTTAAACATTTCTCAAGCATTTTATCCCTATCTCATCTGAA 60
DB 35 ATTTAATTTTCTAATTAACATTTCTCAAGCATTTTATCCCTATCTCATCTGAA 94
QY 61 TTTTAAATAATTAACATTAGTATTAGAAAACTAGAAAAAGATNATGACATATTAA 120
DB 95 TTTTAAATAATTAACATTAGTATTAGAAAACTAGAAAAAGATNATGACATATTAA 154
QY 121 ACTTACATGAAAAAGAAAAATTATACAAAGACTGAGAACTTATTAATTTGAATGAGA 180
DB 155 ACTTACATGAAAAAGAAAAATTATACAAAGACTGAGAACTTATTAATTTGAATGAGA 214
QY 181 TTTAATTTTGAATAATCTGCAAGCACTTATTTGTCATTTCTTATATGATGG 240
DB 215 TTTAATTTTGAATAATCTGCAAGCACTTATTTGTCATTTCTTATATGATGG 274
QY 241 TGTGTTTGACTAATCACTGATTTTCAATAAGAAACCATGTTAAAAATTTTAT 300
DB 275 TGTGTTTGACTAATCACTGATTTTCAATAAGAAACCATGTTAAAAATTTTAT 334
QY 301 TTTTAAATAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTGAATTTGGGAGA 360
DB 335 TTTTAAATAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTGAATTTGGGAGA 394
QY 361 AATTAACCTGTTTCATGATGCAATGCAAAATTTTATGATTTTATCTCCTAATTTT 420
DB 395 AATTAACCTGTTTCATGATGCAATGCAAAATTTTATGATTTTATCTCCTAATTTT 454
QY 421 AAGAACTATGAGAAATTTGATTAATGACATGAGTGCACAACTAATTACTGGCCAGCT 480
DB 455 AAGAACTATGAGAAATTTGATTAATGACATGAGTGCACAACTAATTACTGGCCAGCT 514
QY 481 GTTGCACTGTGTTTCTTACTTACTTACTTCCCAAGAAACTCTTAACTGAATCTTCAGC 540
DB 515 GTTGCACTGTGTTTCTTACTTACTTACTTCCCAAGAAACTCTTAACTGAATCTTCAGC 574
QY 541 NGAATAA-CCTTAATATCTTTGTTAGCCAAACAAA--CTTTTGTGTTTACATAGTTC 597
DB 575 AGAATAATCCCTTAATATCTTTGTTAGCCAAACAAAAGCTTTTGTGTTTACATAGTTC 634
QY 598 TTTGGAATTTTACTGTTTCTTAAATTTTATTTGAAATCTCATTTTCCCAACCATATTA 657
DB 635 TTTGGAATTTTACTGTTTCTTAAATTTTATTTTGAATCTCATATTTTACCACCATATTA 694
QY 658 CCTTATTTAATCTTTGTTATGCAAGTGT 686
DB 695 CCATA-TTAACTTTGATGACAGTGT 722

```

RESULT 7  
BU619112 759 bp mRNA linear EST 23-SEP-2002  
LOCUS BU619112  
DEFINITION UI-H-FH1-bfm-1-10-0-UI.s1 NCI\_CGAP\_FH1 Homo sapiens cDNA clone

ACCESSION UI-H-FH1-bfm-1-10-0-UI 3', mRNA sequence.

VERSION BU619112  
KEYWORDS BU619112.1 GI:23285327  
SOURCE EST.

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 759)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 759)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Contact: Robert Strausberg, Ph.D.  
Email: cgapsb@mail.nih.gov

COMMENT Tissue Procurement: James Martin  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bent-soares@uiowa.edu

The following repetitive elements were found in this cDNA  
sequence: 316-342, >AT-richLow\_complexity (matched complement)  
Seg primer: M13 FORWARD

POLYA=Yes.

FEATURES location/Qualifiers

1..759  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-FH1-bfm-1-10-0-UI"  
/tissue\_type="Cell Line"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCCGCGC. The cell line was provided by Dr. James Martin from the University of Iowa."

BASE COUNT 269 a 107 c 97 g 286 t  
ORIGIN

Query Match 92.5%; Score 635.8; DB 13; Length 759;  
Best Local Similarity 97.8%; Pred. No. 8.8e-66;  
Matches 674; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

```

QY 1 ATTTAATTTTCTTAAACATTTCTCAAGCATTTTATCCCTATCTCATCTGAA 60
DB 32 ATTTAATTTTCTAATTAACATTTCTCAAGCATTTTATCCCTATCTCATCTGAA 91
QY 61 TTTTAAATAATTAACATTAGTATTAGAAAACTAGAAAAAGATNATGACATATTAA 120
DB 92 TTTTAAATAATTAACATTAGTATTAGAAAACTAGAAAAAGATNATGACATATTAA 151
QY 121 ACTTACATGAAAAAGAAAAATTATACAAAGACTGAGAACTTATTAATTTGAATGAGA 180
DB 152 ACTTACATGAAAAAGAAAAATTATACAAAGACTGAGAACTTATTAATTTGAATGAGA 211

```

QY 181 TTATTAATTTGAAAACTGATCTGAAAGCAACTTATTTGTCATTAATTTCTTAATGATGG 240  
 DB 212 TTATTAATTTGAAAACTGATCTGAAAGCAACTTATTTGTCATTAATTTCTTAATGATGG 271  
 QY 241 TGTTTTATGACTATACATGATTTTTCATTAATTAAGAAACCATGTATAAATATTTTAT 300  
 DB 272 TGTTTTATGACTATACATGATTTTTCATTAATTAAGAAACCATGTATAAATATTTTAT 331  
 QY 301 TTTAAAAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTTGAATTTGGAGA 360  
 DB 332 TTTAAAAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTTGAATTTGGAGA 391  
 QY 361 AATATCTGTTTCTGATGATGAAAGTCAAAATTTTGAATTTTATCTGATTTTATTTT 420  
 DB 392 AATATCTGTTTCTGATGATGAAAGTCAAAATTTTGAATTTTATCTGATTTTATTTT 451  
 QY 421 AAGAACTATTAAGAAATTTGATTAATGAATGAAGTCAACACTTAATTTACTGGCCAGCT 480  
 DB 452 AAGAACTATTAAGAAATTTGATTAATGAATGAAGTCAACACTTAATTTACTGGCCAGCT 511  
 QY 481 GTTGCAATTTGTTTCTTACTTACTTACTTCCCAAGAAACTCTTAACTGAATCTTCAGC 540  
 DB 512 GTTGCAATTTGTTTCTTACTTACTTACTTCCCAAGAAACTCTTAACTGAATCTTCAGC 571  
 QY 541 NGAAATTA-CCTTAATATACTTTTGTAGCCAAACAAA--CTTTTGTGTTACATAGTTC 597  
 DB 572 AGAATTAATCCTTAATATACTTTTGTAGCCAAACAAA--CTTTTGTGTTACATAGTTC 631  
 QY 598 TTTGATTTTATCTGTTCTTATTTTATCTGAAATCTCATTTTCCCGAGACCTAATTA 657  
 DB 632 TTTGATTTTATCTGTTCTTATTTTATCTGAAATCTCATTTTCCCGAGACCTAATTA 691  
 QY 658 CCTATTATTAATCTTTGTTATGACAGATTGT 686  
 DB 692 CCATTA-TTAATCTTTGATTAATGACAGATTGT 719

RESULT 8  
 A1675901 714 bp mRNA linear EST 17-DEC-1999  
 LOCUS uc06h04.x1 NCI CGAP P28 Homo sapiens cDNA clone IMAGE:2314423 3'  
 DEFINITION similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR  
 (HUMAN); mRNA sequence.

ACCESSION A1675901  
 VERSION A1675901.1 GI:4876381  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.  
 TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 COMMENT Tumor Gene Index  
 Unpublished

JOURNAL CONTACT: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaapb-r@mail.nih.gov  
 Tissue procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmerit-Buck, M.D., Ph.D.  
 cDNA library preparation: M. Bento Soares, Ph.D.  
 cDNA library arrayed by: Greg Lennon, Ph.D.  
 DNA sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www.bio.lnl.gov/bbrp/image/image.html  
 Insert Length: 1429 Sca Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 439.  
 Location/Qualifiers

FEATURES

source 1..714  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"

/db\_xref="taxon:9606"  
 /clone="IMAGE:2314423"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_P28"  
 /note="Organ: prostate; Vector: pT7AD-Pac (Pharmacia)  
 with a modified polylinker; Plasmid DNA from the  
 normalized library NCI CGAP P22 was prepared, and as  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clonids  
 985608-986759, 110192-110199, and 121928-122015)."  
 Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 257 a 98 c 92 g 263 t 4 others  
 ORIGIN  
 Query Match 92.5%; Score 635.4; DB 9; Length 714;  
 Best Local Similarity 97.1%; Pred. No. 1e-65;  
 Matches 664; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

QY 1 ATTTTAATTTTCTAATTAACATTTCTTCAAGCATTTATTTATCTATATCTCACTGAA 60  
 DB 12 ATTTTAATTTTCTAATTAACATTTCTTCAAGCATTTATTTATCTATATCTCACTGAA 71  
 QY 61 TTTTAAAAATTAACATTTATTTAGAAAATCTAGAAAATTAATGAGATTAATTA 120  
 DB 72 TTTTAAAAATTAACATTTATTTAGAAAATCTAGAAAATTAATGAGATTAATTA 131  
 QY 121 ACTTACATGAAAAAGAAAATTTATTAACAAGACTGAGAACTGATTAATTAATGAATGAGA 180  
 DB 132 ACTTACATGAAAAAGAAAATTTATTAACAAGACTGAGAACTGATTAATTAATGAATGAGA 191  
 QY 181 TTATTAATTTGAAAACTGATCTGAAAGCAACTTATTTGTCATTAATTTCTTAATGATGG 240  
 DB 192 TTATTAATTTGAAAACTGATCTGAAAGCAACTTATTTGTCATTAATTTCTTAATGATGG 251  
 QY 241 TGTTTTATGACTATACATGATTTTTCATTAATTAAGAAACCATGTATAAATATTTTAT 300  
 DB 252 TGTTTTATGACTATACATGATTTTTCATTAATTAAGAAACCATGTATAAATATTTTAT 311  
 QY 301 TTTAAAAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTTGAATTTGGAGA 360  
 DB 312 TTTAAAAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTTGAATTTGGAGA 371  
 QY 361 AATATCTGTTTCTGATGATGAAAGTCAAAATTTTGAATTTTATCTGATTTTATTTT 420  
 DB 372 AATATCTGTTTCTGATGATGAAAGTCAAAATTTTGAATTTTATCTGATTTTATTTT 431  
 QY 421 AAGAACTATTAAGAAATTTGATTAATGAATGAAGTCAACACTTAATTTACTGGCCAGCT 480  
 DB 432 AAGAACTATTAAGAAATTTGATTAATGAATGAAGTCAACACTTAATTTACTGGCCAGCT 491  
 QY 481 GTTGCAATTTGTTTCTTACTTACTTACTTCCCAAGAAACTCTTAACTGAATCTTCAGC 540  
 DB 492 GTTGCAATTTGTTTCTTACTTACTTACTTCCCAAGAAACTCTTAACTGAATCTTCAGC 551  
 QY 541 NGAAATTA-CCTTAATATACTTTGTTAGCCAAACAAA--CTTTTGTGTTACATAGTTC 599  
 DB 552 AGAATTAATCCTTAATATACTTTGTTAGCCAAACAAA--CTTTTGTGTTACATAGTTC 611  
 QY 600 TGTATTTTATGCTTTCTAATTTTATTTTGAATTTTATTTTCCCAAGCATTAATTTAC 659  
 DB 612 TGTATTTTATGCTTTCTAATTTTATTTTGAATTTTATTTTCCCAAGCATTAATTTAC 671  
 QY 660 CTATTACCTTTGTTAGCAGCT 683  
 DB 672 ATA-TTAATCTTTGATTAATGACAGCT 694

RESULT 9  
 BM985376

LOCUS BM985376 756 bp mRNA linear EST 20-FEB-2003  
 DEFINITION UI-CF-EC1-acg-p-23-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone  
 ACCESSION UI-CF-EC1-acg-p-23-0-UI.3', mRNA sequence.  
 BM985376  
 VERSION BM985376.1 GI:19611803  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 756)  
 Author(s) Ronald M.F., Lennon, G. and Soares, M.B.  
 Title Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul.mccray@iowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.reagen.com) or from Open Biosystems  
 (www.openbiosystems.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 316-342, >AT rich#Low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes

## FEATURES

Source

Location/Qualifiers

1..756  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-EC1-acg-p-23-0-UI"  
 /tissue\_type="Lung"  
 /dev\_stage="Adult and Fetal"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_1ib="UI-CF-EC1"  
 /note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
 UI-CF-EC1 is a normalized cDNA library containing the  
 following tissue(s): Normal lung from adult and from fetal  
 day 64, day 87, week 19 and week 42. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pRTT3-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 AAGTCTTAC.  
 TAG\_L1B=UI-CF-EC1  
 TAG\_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371  
 and 380-383  
 TAG\_SEQ=AAGTCTTAC"  
 BASE COUNT 268 a 107 c 97 g 283 t 1 others  
 ORIGIN

Query Match 92.5%; Score 635.2; DB 12; Length 756;  
 Best Local Similarity 97.8%; Pred. No. 1e-65;  
 Matches 673; Conservative 0; Mismatches 12; Indels 3; Gaps 3;  
 1 ATTTAATTTTCTAATTAACATTTCTCAAGCATTTATTTTCTATATCTCACTGAA 60  
 J|||||

Db 32 ATTTAATTTTCTAATTAACATTTCTCAAGCATTTATTTTCTATATCTCACTGAA 91  
 Qy 61 TTTTANANAATPACCTTGTATTTAGAAAAAAGATTAATGACATATTTA 120  
 Db 92 TTTTAAAGAAATPACCTTGTATTTAGAAAAAAGATTAATGACATATTTA 151  
 Qy 121 ACTTACATGAAAAAGAAAAATTTATACAAAGAGTGAAGCTTTAATTAATGAATGGA 180  
 Db 152 ACTTACATGAAAAAGAAAAATTTATACAAAGAGTGAAGCTTTAATTAATGAATGGA 211  
 Qy 181 TTAATATTTGAAAACTGATCTGAAAGCAAACTTTATTTGTCATTAATTAATGATG 240  
 Db 212 TTATATTTGAAAACTGATCTGAAAGCAAACTTTATTTGTCATTAATTAATGATG 271  
 Qy 241 TGTTTATGACTAATPACCTGATTTTGAATTAAGAAACCATGTTAAAAATATTTAT 300  
 Db 272 TGTTTATGACTAATPACCTGATTTTGAATTAAGAAACCATGTTAAAAATATTTAT 331  
 Qy 301 TTTAAAAATPAGCTGTGTGACGCTGATCATATTTCTTTATTTGATTTGGGAAGA 360  
 Db 332 TTTAAAAATPAGCTGTGTGACGCTGATCATATTTCTTTATTTGATTTGGGAAGA 391  
 Qy 361 AAATACGTGTTCTGATGACATGAATGCAAAATTTTATGATTTTAAATCTGCTAATTT 420  
 Db 392 AAATACGTGTTCTGATGACATGAATGCAAAATTTTATGATTTTAAATCTGCTAATTT 451  
 Qy 421 AAGAACTTTAGAAATTTGATTAATGACATGAAGTGAACAACTTAATTTAGTCCAGCT 480  
 Db 452 AAGAACTTTAGAAATTTGATTAATGACATGAAGTGAACAACTTAATTTAGTCCAGCT 511  
 Qy 481 GTTGCAATGTTGTTCTTACTAGTTCTCCCAAGAAACCTTAACCTGAATCTTCAAGC 540  
 Db 512 GTTGCAATGTTGTTCTTACTAGTTCTCCCAAGAAACCTTAACCTGAATCTTCAAGC 571  
 Qy 541 NGAAATTA-CCTTAATATATCTTTGTTAGCCAAACAAA-CTTTTGTGTTACATAGTTCT 598  
 Db 572 AGAATAATCCTTAATATATCTTTGTTAGCCAAACAAA-CTTTTGTGTTACATAGTTCT 631  
 Qy 599 TTGATTTTACTGTTCTTAATTTTATTTCTGAACCTCATTTTCCCAAGCAATATTAC 658  
 Db 632 TTGATTTTACTGTTCTTAATTTTATTTCTGAACCTCATTTTCCCAAGCAATATTAC 691  
 Qy 659 CCTATTAATCTTTGTTAGCAGTTGT 686  
 Db 692 CATTAATCTTTGTTAGCAGTTGT 718

RESULT 10  
 BUE27147 762 bp mRNA linear EST 23-SEP-2002  
 LOCUS BUE27147  
 DEFINITION UI-H-FG0-bct-i-12-0-UI.s1 NCI\_CGAP\_EN1\_2 Homo sapiens cDNA clone  
 UI-H-FG0-bct-i-12-0-UI.3', mRNA sequence.  
 ACCESSION BUE27147  
 VERSION BUE27147.1 GI:23293361  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 762)  
 Author(s) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 Title National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: James Martin  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@iowa.edu  
 The following repetitive elements were found in this cDNA



sequence: 1-43, >AT rich#Low complexity (matched compliment)  
309-335, >AT rich#Low complexity (matched compliment)  
Seq primer: N13 FORWARD  
POLYATes.

## FEATURES

source

Location/Qualifiers

1..762  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-FG0-Dec-1-12-0-UI"  
/tissue\_type="Enchondroma cell line"  
/dev\_stages="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP EN1.2"  
/note="Organ: Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP EN1.2 is a cDNA library containing the following tissue(s): Enchondroma cell line (2 cell lines). The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (GT)18 tail. The sequence tag for this library is CGGTCACTC. The cell lines was provided by Dr James Martin from University of Iowa.  
TAG LIB=UI-H-FG0  
TAG TISSUE=Enchondroma cell line (MLX of EN1 and EN2)  
TAG\_SEQ=CGGTCACTC"

BASE COUNT 269 a 109 c 97 g 286 t 1 others  
ORIGIN

Query Match 92.5%; Score 635.2; DB 13; Length 762;  
Best Local Similarity 97.8%; Pred. No. 1e-65;

Matches 673; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

1 ATTTAATTTCTAATTAACATCTTCTCAAGCATTTATTCCTATATCTCACTGAA 60  
25 ATTTAATTTCTAATTAACATCTTCTCAAGCATTTATTCCTATATCTCACTGAA 84  
61 TTTTAAATAATTAACATCTTCTCAAGCATTTATTCCTATATCTCACTGAA 120  
85 TTTTAAATAATTAACATCTTCTCAAGCATTTATTCCTATATCTCACTGAA 144  
121 ACTTACATGAAGAAATTAATTAACAAGCACTGAGACGTTATTAATTGAATGAGA 180  
145 ACTTACATGAAGAAATTAATTAACAAGCACTGAGACGTTATTAATTGAATGAGA 204  
181 TTAATAATTTGAAGAAATTAATTAACAAGCACTGAGACGTTATTAATTGAATGAGA 240  
205 TTAATAATTTGAAGAAATTAATTAACAAGCACTGAGACGTTATTAATTGAATGAGA 264  
241 TGTATTATGACATTAATCACTGATTTTCAATTAAGAAACCATGTTAAATATTTTAT 300  
265 TGTATTATGACATTAATCACTGATTTTCAATTAAGAAACCATGTTAAATATTTTAT 324  
301 TTTTAAATAATTAACATCTTCTCAAGCATTTATTCCTATATCTCACTGAA 360  
335 TTTTAAATAATTAACATCTTCTCAAGCATTTATTCCTATATCTCACTGAA 384  
361 AAATACTGTTCTGATAGCATGAATGCAAAATTTTATAGATTTTATCTGCTAATTTT 420  
385 AAATACTGTTCTGATAGCATGAATGCAAAATTTTATAGATTTTATCTGCTAATTTT 444  
421 AAGAATCTATGGAATTAATTAATGAACATGAAGTGCACAACTAATTAATCTGGCAGCT 480  
445 AAGAATCTATGGAATTAATTAATGAACATGAAGTGCACAACTAATTAATCTGGCAGCT 504  
481 GTTGCAATGTTCTTCTAGTCTCCCAAGGAAACCTTAACTGAATCTTCAGC 540

Db 505 GTTGCAATGTTCTTCTAGTCTCCCAAGGAAACCTTAACTGAATCTTCAGC 564  
Qy 541 NGAATTA-CTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 598  
Db 565 AAGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 624  
Qy 599 TTTGATTTTACTGTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 658  
Db 625 TTTGATTTTACTGTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 684  
Qy 659 CCTATTAACTTTGTTATGACAGTTGT 686  
Db 685 CAT-ATTAATTTGTTATGACAGTTGT 711

RESULT 11  
LOCUS A1826287 770 bp mRNA linear EST 21-DEC-1999  
DEFINITION wk3f07.x1 NCI CGAP Pr22 Homo sapiens cDNA clone IMAGE:2417221.3.  
similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR  
(HUMAN); contains element MRR28 repetitive element ;, mRNA sequence.

ACCESSION A1826287  
VERSION A1826287.1 GI:5446958  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-rc@mail.nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnlnl.gov/bbrp/image/image.html  
Insert Length: 1454 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence sloop: 404.

## FEATURES

source

Location/Qualifiers

1..770  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2417221"  
/sex="male"  
/tissue\_type="normal prostate"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Pr22"  
/note="Organ: prostate; Vector: pT73-Pac (Pharmacia)  
with a modified polylinker; 1st strand cDNA was prepared  
from normal prostate bulk tissue, and was then primed with  
a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT73 vector. Library is normalized, and was  
constructed by Bento Soares and M. Facima Bonaldo. "

BASE COUNT 271 a 111 c 99 g 285 t 4 others  
ORIGIN

Query Match 92.3%; Score 634.4; DB 9; Length 770;  
Best Local Similarity 97.5%; Pred. No. 1.3e-65;

Matches 671; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

1 ATTTAATTTCTAATTAACATCTTCTCAAGCATTTATTCCTATATCTCACTGAA 60

```

Db      38 ATTTTATTTTCATTAACATCTCTGCAAGCATTTATTTATCTATATCTCACTGAA 97
Qy      61 TTTTAAANAATAACATTAGTATTGAAAACTAGAAAAAGATTTATGCAATATTAA 120
Db      98 TTTTAAAGAAATTAACATTAGTATTGAAAACTAGAAAAAGATTTATGCAATATTAA 157
Qy      121 ACTTACATGAAAAAGAAAAATTTAACAAGGCTGAAACGTTATTAATTTGAAAAAGAA 180
Db      158 ACTTACATGAAAAAGAAAAATTTAACAAGGCTGAAACGTTATTAATTTGAAAAAGAA 217
Qy      181 TTAATATTGAAAAAGCTGCACTGAAAGCAAACTTTATGTTCAATTAATCTTAATGATG 240
Db      218 TTAATATTGAAAAAGCTGCACTGAAAGCAAACTTTATGTTCAATTAATCTTAATGATG 277
Qy      241 TGTATTATGACTATATACATGATTTTTCATTAAGAAACCCATGTTAAAAATTTTAT 300
Db      278 TGTATTATGACTATATACATGATTTTTCATTAAGAAACCCATGTTAAAAATTTTAT 337
Qy      301 TTTTAAAAATGAGCGTGTGTCAGAGCTGTGATCATATTTCTTTATTTGATTTGGAGAA 360
Db      338 TTTTAAAAATGAGCGTGTGTCAGAGCTGTGATCATATTTCTTTATTTGATTTGGAGAA 397
Qy      361 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTATCTCNCATAATTT 420
Db      398 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTATCTCNCATAATTT 457
Qy      421 AAGAACTATTGGAATTTGATTTATGATGAAAGTGCACACACTAATTACTGGCCAGCT 480
Db      458 AAGAACTATTGGAATTTGATTTATGATGAAAGTGCACACACTAATTACTGGCCAGCT 517
Qy      481 GTTGCGATGTGTTCTTACTAGTCTGCCAAGAAAACTTTAAGCAATCTTCAGC 540
Db      518 GTTGCGATGTGTTCTTACTAGTCTGCCAAGAAAACTTTAAGCAATCTTCAGC 577
Qy      541 NGAATTA-CCTTAATATATCTTGTGTAGCCAAACAAA-CTTTTGTGTACATAGTCT 598
Db      578 AGAATATCTTAAATATATCTTGTGTAGCCAAACAAAAGCTTNTTGTGTACATAGTCT 637
Qy      599 TTGGATTTTACTGTTCTTAATTTTATTTCTGAAACCTCATTTCCTCCAGCCATAATTAC 658
Db      638 TTGGATTTTACTGTTCTTAATTTTATTTCTGAAACCTCATTTCCTCCAGCCATAATTAC 697
Qy      659 CCTATTACTTGTGTATGACAGCTGT 686
Db      698 CATTA-TTAACCTNTGATAGCAGCTGT 724

RESULT 12
BM968746      721 bp      mRNA      linear      EST 20-FEB-2003
LOCUS      UI-CF-DUI-aam-a-03-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
DEFINITION      UI-CF-DUI-aam-a-03-0-UI 3', mRNA sequence.
ACCESSION      BM968746
VERSION      BM968746.1 GI:19586333
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 721)
Normalisation and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL      97044477
MEDLINE      8889548
PUBMED
COMMENT      Mccray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu

```

Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com) or from Open Biosystems  
 (www.openbiosystems.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 315-341, >AT-richlow\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=yes.

FEATURES  
 source  
 1..721  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-DUI-aam-a-03-0-UI"  
 /issue\_type="Primary Lung Epithelial Cells"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-CF-DUI"  
 /note="Organ: Lung; Vector: pTR73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-DUI is a normalized cDNA library containing the  
 following tissue(s): Primary Lung Epithelial Cells The  
 library was constructed according to Bonaldo, Lennon and  
 Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an EcoR I adaptor, digested with Not I, and cloned  
 directionally into pTR73-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dT)18 tail. The sequence tag for this  
 library is GGCTGTAGGC.  
 TAG LIB=UI-CF-DUI  
 TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368  
 TAG\_SEQ=GGCTGTAGGC"

BASE COUNT 257 a 97 c 91 g 276 t

ORIGIN

Query Match 92.0%; Score 632; DB 12; Length 721;  
 Best Local Similarity 97.5%; Pred. No. 2,5e-65;  
 Matches 671; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

```

Qy      1 ATTTTATTTTCATTAACATCTCTGCAAGCATTTATTTATCTATATCTCACTGAA 60
Db      31 ATTTTATTTTCATTAACATCTCTGCAAGCATTTATTTATCTATATCTCACTGAA 90
Qy      61 TTTTAAANAATAACATTAGTATTGAAAACTAGAAAAAGATTTATGCAATATTAA 120
Db      91 TTTTAAAGAAATTAACATTAGTATTGAAAACTAGAAAAAGATTTATGCAATATTAA 150
Qy      121 ACTTACATGAAAAAGAAAAATTTAACAAGGCTGAAACGTTATTAATTTGAAAAAGAA 180
Db      151 ACTTACATGAAAAAGAAAAATTTAACAAGGCTGAAACGTTATTAATTTGAAAAAGAA 210
Qy      181 TTAATATTGAAAAAGCTGCACTGAAAGCAAACTTTATGTTCAATTAATCTTAATGATG 240
Db      211 TTAATATTGAAAAAGCTGCACTGAAAGCAAACTTTATGTTCAATTAATCTTAATGATG 270
Qy      241 TGTATTATGACTATATACATGATTTTTCATTAAGAAACCCATGTTAAAAATTTTAT 300
Db      271 TGTATTATGACTATATACATGATTTTTCATTAAGAAACCCATGTTAAAAATTTTAT 330
Qy      301 TTTTAAAAATGAGCGTGTGTCAGAGCTGTGATCATATTTCTTTATTTGATTTGGAGAA 360
Db      331 TTTTAAAAATGAGCGTGTGTCAGAGCTGTGATCATATTTCTTTATTTGATTTGGAGAA 390
Qy      361 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTATCTCNCATAATTT 420
Db      391 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTATGATTTTATCTCNCATAATTT 450

```

QY 421 AAGAACTATTGGAATGATTAATGACATGAGTGCACACATTAATCTGCGAGCT 480  
 DB 451 AAGAACTATTGGAATGATTAATGACATGAGTGCACACATTAATCTGCGAGCT 510  
 QY 481 GTTGGCATTGTGTTCTTACTAGTTCCTCCAGGAAAACCTTTAACTGAATCTTTCAGC 540  
 DB 511 GTTGGCATTGTGTTCTTACTAGTTCCTCCAGGAAAACCTTTAACTGAATCTTTCAGC 570  
 QY 541 NGAATTA-CCTTAATTAATCTTGTAGCCAAACAAA-CTTTTGTGTACATAGTTCT 598  
 DB 571 AGAATTAATCTTAAATTAATCTTGTAGCCAAACAAAACCTTTTGTGTACATAGTTCT 630  
 QY 599 TTGGATTTTACTGTTCTTAACTTTTCTGAAAGCTCATTTTCCCGACCATTAATTC 658  
 DB 631 TTGGATTTTACTGTTCTTAACTTTTCTGAAAGCTCATTTTCCCGACCATTAATTC 690  
 QY 659 CCTATTTAACCTTTGTATGACACAGTTGT 686  
 DB 691 CATTA-TTAACCTTTGTATGACACAGTTGT 717  
 RESULT 13  
 CA424156 734 bp mRNA linear EST 07-NOV-2002  
 LOCUS UI-H-FEI-bdv-m-12-0-UI-81 NCI CGAP FEI Homo sapiens cDNA clone  
 DEFINITION UI-H-FEI-bdv-m-12-0-UI 3', mRNA sequence.  
 ACCESSION CA424156  
 VERSION CA424156  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 734)  
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaapb-r@mail.nih.gov  
 Tissue Procurement: James Martin  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-48, >AT richLow complexity (matched complement)  
 314-340, >AT richLow complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYAref8.  
 FEATURES  
 source Location/Qualifiers  
 1..734  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FEI-bdv-m-12-0-UI"  
 /csize\_type="Cell lines"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI CGAP FEI"  
 /note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia)  
 with a modified polylinker; Site\_1: EcoR I; Site\_2: Not  
 I; NCI CGAP FEI is a normalized cDNA library derived from  
 a pool of mRNA obtained from 3 cell lines from grade II  
 chondrosarcoma tissues. The library was constructed  
 according to Bonaldo, Lennon and Soares, Genome Research,  
 6:791-806, 1996. First strand cDNA synthesis was primed  
 with an oligo-dT primer containing a Not I site. Double  
 stranded cDNA was ligated to an EcoR I adaptor, digested  
 with Not I, and cloned directionally into pT73-Pac  
 vector. The oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The  
 sequence tag for this library is CGCTACGGAC. The cell  
 lines were provided by Dr James Martin from the University  
 of Iowa.  
 TAG\_LIB=UI-H-FEI  
 TAG\_TISSUE=Human grade 2 chondrosarcoma cell line pool  
 TAG\_SEQ=CGCTACGGAC"  
 BASE COUNT 261 a 101 c 94 g 277 t 1 others  
 ORIGIN  
 Query Match 92.0%; Score 632; DB 14; Length 734;  
 Best Local Similarity 97.5%; Pred. No. 2-5e-65;  
 Matches 671; Conservative 0; Mismatches 14; Indels 3; Gaps 3;  
 1 ATTTAATTTTCTAATAACATTTCTTCAAGCATTTTATTTATCTATCTCACTGA 60  
 DB 30 ATTTAATTTTCTAATAACATTTCTTCAAGCATTTTATTTATCTATCTCACTGA 89  
 QY 61 TTTTAAATAATACATTAATTAAGAAAACCTAGAAAATAAGATTAATTAATTA 120  
 DB 90 TTTTAAATAATACATTAATTAAGAAAACCTAGAAAATAAGATTAATTAATTA 149  
 QY 121 ACTTACATGAAAAGAAAATTTATACAAAGACTAGAACCTTATTAATTAATGA 180  
 DB 150 ACTTACATGAAAAGAAAATTTATACAAAGACTAGAACCTTATTAATTAATGA 209  
 QY 181 TTATATTTTGAAGAACTGATCTGAAAGCAACTTTATTTGTTCAATTAATTA 240  
 DB 210 TTATATTTTGAAGAACTGATCTGAAAGCAACTTTATTTGTTCAATTAATTA 269  
 QY 241 TGTATTATGACATTAATACATGATTTTCAATTAAGAAACCATTAATAATTTTAT 300  
 DB 270 TGTATTATGACATTAATACATGATTTTCAATTAAGAAACCATTAATAATTTTAT 329  
 QY 301 TTTTAAATAATACATTAATTAAGAAACCTGATCATATTTCTTTATTTGATTTG 360  
 DB 330 TTTTAAATAATACATTAATTAAGAAACCTGATCATATTTCTTTATTTGATTTG 389  
 QY 361 AATATCTGTTTCTGATGATGAATAATGCAAAATTTTATGATTTTATCTCTAATTT 420  
 DB 390 AATATCTGTTTCTGATGATGAATAATGCAAAATTTTATGATTTTATCTCTAATTT 449  
 QY 421 AAGAACTATTGGAATGATTAATGACATGAGTGCACACATTAATCTGCGAGCT 480  
 DB 450 AAGAACTATTGGAATGATTAATGACATGAGTGCACACATTAATCTGCGAGCT 509  
 QY 481 GTTGGCATTGTGTTCTTACTAGTTCCTCCAGGAAAACCTTTAACTGAATCTTTCAGC 540  
 DB 510 GTTGGCATTGTGTTCTTACTAGTTCCTCCAGGAAAACCTTTAACTGAATCTTTCAGC 569  
 QY 541 NGAATTA-CCTTAATTAATCTTGTAGCCAAACAAA-CTTTTGTGTACATAGTTCT 598  
 DB 570 AGAATTAATCTTAAATTAATCTTGTAGCCAAACAAAACCTTTTGTGTACATAGTTCT 629  
 QY 599 TTGGATTTTACTGTTCTTAACTTTTCTGAAAGCTCATTTTCCCGACCATTAATTC 658  
 DB 630 TTGGATTTTACTGTTCTTAACTTTTCTGAAAGCTCATTTTCCCGACCATTAATTC 689  
 QY 659 CCTATTTAACCTTTGTATGACACAGTTGT 686  
 DB 690 CATTA-TTAACCTTTGTATGACACAGTTGT 716  
 RESULT 14  
 AM052045 786 bp mRNA linear EST 20-SEP-1999  
 LOCUS wX5905.x1 NCI CGAP Kid1 Homo sapiens cDNA clone IMAGE:2544704.3  
 DEFINITION similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR  
 (HUMAN); mRNA sequence.  
 ACCESSION AM052045  
 VERSION AM052045  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)



sequence tag for this library is GCTCAAGCT.

TAG\_LIB=H-BD1  
TAG\_TISSUE=chondrosarcoma  
TAG\_SEQ=CGTCAAGCTBASE COUNT 266 a 105 c 96 g 281 t 2 others  
ORIGIN

Query Match 91.7%; Score 630.2; DB 12; Length 750;

Best Local Similarity 96.4%; Pred. No. 4e-65;

Matches 663; Conservative 0; Mismatches 23; Indels 2; Gaps 2;

```
QY 1 ATTTAATTTTCTAATTAACATTTCTCTGAAGCATTATTTATCTTATATCTCACTGAA 60
   |||||
Db 32 ATTTAATTTTCTAATTAACATTTCTCTGAAGCATTATTTATCTTATATCTCACTGAA 91
   |||||
QY 61 TTTTAAATAATAACATTAGTATTAGAAAACTAGAAAAAGATNAATGACATTAATTA 120
   |||||
Db 92 TTTTAAATAATAACATTAGTATTAGAAAACTAGAAAAAGATNAATGACATTAATTA 151
   |||||
QY 121 ACTTACATGAAAAAGAAAAATTATACAAAGACTGAAACGTTATAATTGAAATGAGA 180
   |||||
Db 152 ACTTACATGAAAAAGAAAAATTATACAAAGACTGAAACGTTATAATTGAAATGAGA 211
   |||||
QY 181 TTATATTTTGAAGAACTGCATCTGAAGCAAACTTTATGTTCAATTATCTTAATGATG 240
   |||||
Db 212 TTATATTTTGAAGAACTGCATCTGAAGCAAACTTTATGTTCAATTATCTTAATGATG 271
   |||||
QY 241 TGTATTATGACTAATACACTGATTTTTCATTAAGAAACCATGTTAAAAATATTTTAT 300
   |||||
Db 272 TGTATTATGACTAATACACTGATTTTTCATTAAGAAACCATGTTAAAAATATTTTAT 331
   |||||
QY 301 TTTTAAAAATAGCCCTGTGTCTCAAGCTCTGATCATATTCTTTATTTTGGATTTGGAGA 360
   |||||
Db 332 TTTTAAAAATAGCCCTGTGTCTCAAGCTCTGATCATATTCTTTATTTTGGATTTGGAGA 391
   |||||
QY 361 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTATGTTTATCTCNCCTAATTTT 420
   |||||
Db 392 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTATGTTTATCTCNCCTAATTTT 451
   |||||
QY 421 AAGAACTATGAGAAATTGATTAAATGACATGAAGTGCACACACTAATTACTGGCCAGCT 480
   |||||
Db 452 AAGAACTATGAGAAATTGATTAAATGACATGAAGTGCACACACTAATTACTGGCCAGCT 511
   |||||
QY 481 GTTGGCATTTGTTTCTTACTTACTTCTCCCAAGGAAAACTTTAACTGATCTTCAGC 540
   |||||
Db 512 GTTGGCATTTGTTTCTTACTTACTTCTCCCAAGGAAAACTTTAACTGATCTTCAGC 571
   |||||
QY 541 NGAATTA - CCTTAATATATCTTGTAGCAACAAA - CTTTGTGTTACATAGTCT 598
   |||||
Db 572 AAGAAATATCTTAAATATATCTTGTAGCAACAAA - CTTTGTGTTACATAGTCT 631
   |||||
QY 599 TTGATTTTACTGTTCTTAAATTTTATCTGAAACTCATTTTCCCGAGCCATTAATTA 658
   |||||
Db 632 TTGATTTTACTGTTCTTAAATTTTATCTGAAACTCATTTTCCCGAGCCATTAATTA 691
   |||||
QY 659 CCTAATTAACCTTTGTTATGACAGTTGT 686
   |||||
Db 692 CCATATTAACCTTTGTTATGACAGTTGT 719
   |||||
```

Search completed: November 27, 2003, 12:28:12  
Job time : 1852.59 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:01:24 ; Search time 225.747 Seconds  
(without alignments)  
8215.019 Million cell updates/sec

Title: US-09-835-992A-20

Perfect score: 687

Sequence: 1 attctatctctctcaaac.....cttgatcgcacagctgtc 687

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	683	99.4	687 20	AAX40095 Gastric cancer ass
2	636.8	92.7	2663 20	AAX40095 Human liver cell s
3	611.8	89.1	714 20	AAX40094 Gastric cancer ass
4	467.8	68.1	486 24	ABV87267 Human colon cancer
5	411.8	59.9	843 20	AAX40097 Gastric cancer ass
6	300.8	43.8	351 16	AAT19975 Human gene signatu
7	261	38.0	262 22	AAH57293 Human liver specif
8	161	23.4	173 25	ABX83093 Corn ear-derived p

C	9	126.2	18.4	263	15	AA076489	Human genome fragm
C	10	103.6	15.1	397	25	ABX40797	Bovine EST associa
C	11	85.8	12.5	2152	24	AB199762	Mouse ischaemic co
C	12	71.6	10.4	8056	25	AB170246	Haematopoietic cel
C	13	70	10.2	8056	25	AB210246	Haematopoietic cel
C	14	66.8	9.7	7676	24	AB170409	Chemically treated
C	15	66.8	9.7	7676	24	AB134598	Human metastasis a
C	16	66.8	9.7	7676	25	AB210100	Haematopoietic cel
C	17	66.4	9.7	5852	12	AA011710	Dicystostellum plas
C	18	65.4	9.5	10048	24	AB170313	Chemically treated
C	19	65.4	9.5	10048	24	AA61251	Human gene regulat
C	20	65.2	9.5	8056	25	AB210100	Haematopoietic cel
C	21	65	9.5	113515	24	AB134174	Human immune syste
C	22	63.4	9.2	883	22	AA115210	Human breast cance
C	23	63	9.2	6106	22	AA546429	Tumour suppressor
C	24	63	9.2	6106	22	ABX40031	Human chemically p
C	25	63	9.2	6106	24	AB133472	Human immune syste
C	26	63	9.2	13131	24	AB192249	Chemically treated
C	27	61.8	9.0	291	25	ABX48529	Bovine EST associa
C	28	60.8	8.9	6699	24	AA631321	Chemically pretrea
C	29	60.4	8.8	6159	22	AA546370	Tumour suppressor
C	30	60.4	8.8	6159	22	ABN80097	Human chemically m
C	31	60.4	8.8	6944	24	ABX34026	Human DNA for stag
C	32	60.2	8.8	6509	24	AB132226	Human immune syste
C	33	60.2	8.8	6912	24	ABK28371	DNA transcription
C	34	59.8	8.7	7459	24	ABK31383	Signal transductio
C	35	59.8	8.7	17721	24	AB133729	Human immune syste
C	36	59.6	8.7	6175	24	AA028372	Human chemically t
C	37	59.6	8.7	10467	22	ABK28453	DNA transcription
C	38	59.4	8.6	5815	24	ABX40024	Human novel protei
C	39	59.2	8.6	5815	24	ABX40024	Buchnera sp. genom
C	40	59.2	8.6	640681	24	ABX92787	Buchnera sp. genom
C	41	58.8	8.6	449	25	ABX48095	Chemically treated
C	42	58.8	8.6	6071	24	AB192215	Chemically treated
C	43	58.8	8.6	6071	24	ABD22316	Chemically treated
C	44	58.4	8.5	424	25	ABX46053	Bovine EST associa
C	45	58.2	8.5	5962	24	AB133287	Human immune syste

#### ALIGNMENTS

RESULT 1	
AA40095	
ID	AA40095 standard; DNA; 687 BP.
XX	
AC	AAX40095;
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Gastric cancer associated gene.
XX	
KW	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO9904265-A2.
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998; 98WO-0514679.
XX	
PR	22-JUN-1998; 98US-0102322.
PR	17-JUL-1997; 97US-0896164.
PR	10-OCT-1997; 97US-0061589.
PR	10-OCT-1997; 97US-0061765.
PR	10-OCT-1997; 97US-0948705.
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDW-) LUDWIG INST CANCER RES.
XX	

PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;  
PI Tureci O;  
XX WPI: 1999-132448/11.  
DR  
XX  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
XX  
PS Claim 67, Page 695; 787bp; English.  
XX  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
XX  
SQ Sequence 687 BP; 242 A; 98 C; 85 G; 258 T; 4 other;

Query Match 99.4%; Score 683; DB 20; Length 687;  
Best Local Similarity 100.0%; Pred. No. 4.6e-108;  
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTAATTTCTTAAACCTCTTCAAGCATTTATTTATCTATATCTACGTAA 60  
DB 1 ATTTAATTTCTTAAACCTCTTCAAGCATTTATTTATCTATATCTACGTAA 60

QY 61 TTTTAAATAATTAACATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 120  
DB 61 TTTTAAATAATTAACATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 120

QY 121 ACTTACATGAAAAAGAAATTTATTAAGAAAGTGAAGCGTTATTAATGAATGAGA 180  
DB 121 ACTTACATGAAAAAGAAATTTATTAAGAAAGTGAAGCGTTATTAATGAATGAGA 180

QY 181 TTTAATTTGAAAACCTGATCTGAAGCAACTTTATTTGCAATTAATTTAATGATGG 240  
DB 181 TTTAATTTGAAAACCTGATCTGAAGCAACTTTATTTGCAATTAATTTAATGATGG 240

QY 241 TGTTTTATGCTATATACGATTTTCAATTAAGAAAGTAAAGTAAAGTAAAGTAA 300  
DB 241 TGTTTTATGCTATATACGATTTTCAATTAAGAAAGTAAAGTAAAGTAAAGTAA 300

QY 301 TTTTAAATAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTGAATGGAGA 360  
DB 301 TTTTAAATAATTAAGCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTGAATGGAGA 360

QY 361 AAATACGTTTCTGATAGCATGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 420  
DB 361 AAATACGTTTCTGATAGCATGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 420

QY 421 AAGACATTTGAGAAATTTGATTAATGACATGAAGTAAAGTAAAGTAAAGTAA 480  
DB 421 AAGACATTTGAGAAATTTGATTAATGACATGAAGTAAAGTAAAGTAAAGTAA 480

QY 481 GTTGACATTTGTTTCTTACTAGTTCTCCAGAGAAACCTTAAACTGAATTTTACG 540  
DB 481 GTTGACATTTGTTTCTTACTAGTTCTCCAGAGAAACCTTAAACTGAATTTTACG 540

QY 541 NGATAATCCTTAATATATCTTTGTTAGCCAAACAAACCTTTTGTTTTACATAGTTCTTT 600  
DB 541 NGATAATCCTTAATATATCTTTGTTAGCCAAACAAACCTTTTGTTTTACATAGTTCTTT 600

QY 601 GGATTTTACTGTTCCATATTTTATTTGAAAGTCAATTTTCCAGACCATATATACC 660  
DB 601 GGATTTTACTGTTCCATATTTTATTTGAAAGTCAATTTTCCAGACCATATATACC 660

QY 661 TATTTAACTTTGTTTATGACAGTTGTT 687  
DB 661 TATTTAACTTTGTTTATGACAGTTGTT 687

RESULT 2  
AAH57501/c  
ID AAH57501 standard; cDNA; 2663 BP.  
XX  
AC AAH57501;  
XX  
DT 10-SEP-2001 (first entry)  
XX  
DE Human liver cell specific cDNA sequence SEQ ID NO:341.  
XX  
KW Human, tissue specific; diagnosis; brain; heart; skeletal muscle;  
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; es;  
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;  
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
XX  
OS Homo sapiens.  
XX  
XX W0200132927-A2.  
XX  
PD 10-MAY-2001.  
XX  
PF 02-NOV-2000; 2000MO-US30396.  
XX  
PR 04-NOV-1999; 99US-0163508.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Sornase T, Sellhammer JD, Watson GA;  
XX  
DR WPI: 2001-291057/30.  
XX  
PT New cell and tissue specific polynucleotides useful for diagnosis,  
PT prognosis or monitoring of treatments for disorders where the gene is  
PT associated with a cancer, immunopathology or neuropathology -  
XX  
PS Claim 1, Page 258-259; 327bp; English.  
XX  
CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
CC sequences (I). (I) can have cytostatic, immunomodulatory and  
CC neuroprotective activities, and can be used in gene therapy. (II) and  
CC proteins (II) encoded by them are used in high throughput screening  
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or  
CC their fragments, immunoglobulins, inhibitors, drug compounds and  
CC pharmaceutical agents. Expression of (I) in a sample indicates the  
CC differentiation of embryonic stem cells into a tissue selected from  
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
CC tissues. (II) and (II) are used to produce an expression profile that  
CC defines a metabolic or developmental process, treatment, condition,  
CC disease or disorder. The gene profile can be used for diagnosis,  
CC prognosis or monitoring of treatments and for investigating a  
CC predisposition to a disorder where the gene is associated with a  
CC cancer, immunopathology or neuropathology.  
XX  
SQ Sequence 2663 BP; 807 A; 472 C; 599 G; 785 T; 0 other;

Query Match 92.7%; Score 636.8; DB 22; Length 2663;  
Best Local Similarity 98.0%; Pred. No. 3.4e-100;  
Matches 674; Conservative 0; Mismatches 11; Indels 3; Gaps 3;

QY 1 ATTTAATTTCTTAAACATTTCTCAAGCATTTATTTATCTATATCTACGTAA 60  
DB 2639 ATTTAATTTCTTAAACATTTCTCAAGCATTTATTTATCTATATCTACGTAA 2580



Oy	6	TTTTAANAATAATPACCTTTNGTATTTGAAAAAATCAGAAAAAAGATTTAAATGCAATATTTAA	120
Db	2579	TTTTTAAGAAATTAACCTTTAGTATTTGAAAAAATCTAGAAAAAAGATTTAAATGCAATATTTAA	2520
Oy	121	ACTTACATGAAAAAGCAAAATTTATPAACTGAAGACGTATTAATTAATGAAATGAGA	180
Db	2519	ACTTTCATGAAAAAGCAAAATTTATPAAAGGACGTATTAATTAATGAAATGAGA	2460
Oy	181	TTATTAATTTGAAAACTGCATCTGAAAAGCAAACTTTATTTGTTCAATTAATCTTAAATGATGG	240
Db	2459	TTATTAATTTGAAAACTGCATCTGAAAAGCAAACTTTATTTGTTCAATTAATCTTAAATGATGG	2400
Oy	241	TGTTTTATGACTAATACACGTATTTTCAATPAGAAACCCATGTTAAAAATATTTTAT	300
Db	2399	TGTTTTATGACTAATACACGTATTTTCAAGAGGAAACCCATGTTAAAAATATTTTAT	2340
Oy	301	TTTAAAAAATAAGCCGTGTCAAGCTCGATCATATTTCTTTAATTTGATTTGGGAAGA	360
Db	2339	TTTAAAAAATAAGCCGTGTGTCAAGCTCGATCATATTTCTTTAATTTGATTTGGGAAGA	2280
Oy	361	AAATACGTGTTCTGTATGACATGAATGCAAAATTTTATGATTTTAACTGCTAAATTTT	420
Db	2279	AAATACGTGTTCTGTATGACATGAATGCAAAATTTTATGATTTTAACTGCTAAATTTT	2220
Oy	421	AAGAACTATTTGAGAAATGATTAATGACATGAATGCAACACTAATTAATTCGGCCAGCT	480
Db	2219	AAGAACTATTTGAGAAATGATTAATGACATGAATGCAACACTAATTAATTCGGCCAGCT	2160
Oy	481	GTTGGCAATGTGTTCTTACTTAAGTCTCCCAAGGAAACCTTAATGCAATCTTCAGC	540
Db	2159	GTTGGCAATGTGTTCTTACTTAAGTCTCCCAAGGAAACCTTAATGCAATCTTCAGC	2100
Oy	541	NGAATPAA-CCTTAAATATACTTTGTTAGCCAAACAAA-CCTTTTGTTCATAGTACT	598
Db	2099	AGAATPAACTTAAATATACTTTGTTAGCCAAACAAAAGCTTTTGTTCATAGTACT	2040
Oy	599	TTGGAATTTTACTGTTCTCTAATTTTATTTCTGAAACTGCAATTTTCCCGACCAATATTAAC	658
Db	2039	TTGGAATTTTACTGTTCTCTAATTTTATTTCTGAAACTGCAATTTTCCCGACCAATATTAAC	1980
Oy	659	CCTATTTTACTTTGTTATGCAAGTGT 686	
Db	1979	CATA-TTAACTTTGTATGACACAGTGT 1953	
RESULT 3			
ID	AAAX40094	standard; DNA; 714 BP.	
AC	AAAX40094;		
XX	02-JUL-1999	(first entry)	
XX	DE	Gastric cancer associated gene.	
XX	KW	Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer; 88.	
XX	OS	Homo sapiens.	
XX	PN	M09904265-A2.	
XX	PD	28-JAN-1999.	
XX	PF	15-JUL-1998; 98WO-US14679.	
XX	PR	22-JUN-1998; 98US-0102322.	
PR	17-OCT-1997;	97US-0896164.	
PR	10-OCT-1997;	97US-0061599.	
PR	10-OCT-1997;	97US-0061765.	
PR	10-OCT-1997;	97US-0948705.	
PR	11-OCT-1997;	97GB-0021697.	

XX	(LUDW-) LUDWIG INST CANCER RES.
PA	
XX	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI	Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
PI	Tureci O;
XX	
DR	WPI, 1999-132448/11.
XX	
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Claim 67; Page 695; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
SQ	Sequence 714 BP, 258 A; 100 C; 92 G; 260 T; 4 other;
	Query Match            89.1%; Score 611.8; DB 20; Length 714;
	Best Local Similarity 97.1%; Pred. No. 6.5e-96;
	Matches 671; Conservative 0; Mismatches 15; Indels 5; Gaps 5
OY	1 ATTTTAATTTTCGTATTAAACATCTTCTGCAAGCATTTATTCCTATATCTCACTGAA 60
DB	15 ATTTTAATTTTCGTATTAAACATCTTCTTCCAAGCATTTATTTATCTTATCTCACTGAA 74
OY	61 TTTTAAAAAATPAACTTAGTATTAGAAAACCTAGAAAAAAGATMAATGCAGATAATTAA 120
DB	75 TTTTAAAGAAATPAACTTAGTATTAGAAAACCTAGAAAAAAGATMAATGCAGATAATTAA 134
OY	121 ACTTACATGAAGAAAAGAAAATTATPACAAGGACTGAGAAGCGTTAAATTGAATGAGA 180
DB	135 ACTTGACATGAAGAAAAGAAAATTATPACAAGGACTGAGAAGCGTTAAATTGAATGAGA 194
OY	181 TTATATATTGAAAACCTGCATCTGAAGAAGAACTTATTTGTTCAATTAATCTTAAATGG 240
DB	195 TTATATATTGAAAACCTGCATCTGAAGAAGAACTTATTTGTTCAATTAATCTTAAATGG 254
OY	241 TGTTTTATGACTAATACACTGATTTTTTCAATPAGAAAACCCATGTTAAAAATATTTTAT 300
DB	255 TGTTTTATGACTAATACACTGATTTTTTCAAGAAAGAAAACCCATGTTAAAAATATTTTAT 314
OY	301 TTTAAAAAATAAGCCGRTGTCAGCTGCATCATATTTCTTTATTTGATTGGGAAGA 360
DB	315 TTTAAAAAATAAGCCGRTGTCAGCTGCATCATATTTCTTTATTTGATTGGGAANA 374
OY	361 AAATACGTGTTCTGATAGCATGAATGCAAAATTTTATGATTTTATATCTCNCTAATTTT 420
DB	375 AAATACGTGTTCTGATAGCATGAATGCAAAATTTTATGATTTTATATCTCACTAATTTT 434
OY	421 AAGAAGCTTTGAGAAATTGATTATATGACATGAAGTGCACACACTAATTTACTGGCCAGCT 480
DB	435 AANAACTATTGAGAAATTGATTATATGACATGAAGTGCACACACTAATTTACTGGCCAGCT 494
OY	481 GTTGSCATGTTGTTTCTTACTAGTTCGCCAA-GGAAAACCTTAAACSTGAATCTTGAG 539
DB	495 GTTGSCATGTTGTTTCTTACTAGTTCGCCAAGGAAAACCTTAAATATGATGATCTTCAG 554
OY	540 CGAATAAT-CCTTAAATATACTTGTGTGTCGCAAAACAAA-CITTTTTGTTTCAATAGTTC 597

Db 555 CAGATATATCCCTTAATATACCTTGTATGACAAACAAAGCCTTTTGTGTTATCATAGTTC 614  
Qy 598 TTT-GGATTTTACGTCTCCCTAATTTTATTCGAAACCTCATTTTCCCGACCAATATT 656  
Db 615 TTTGGGATTTTACGTCTCCCTAATTTTATTCGAAACCTCATTTTCCCGACCAATATT 674  
Qy 657 ACCCATTAATCTTGTATGACAGTGTGTT 687  
Db 675 ACCATA-TTAACTTTGTTTGTGACAGTGTGTT 704

RESULT 4  
ABV87267  
ID ABV87267 standard; cDNA, 486 BP.  
XX  
AC ABV87267;  
XX  
DT 13-DEC-2002 (first entry)  
XX  
DE Human colon cancer related cDNA SEQ ID NO 578.  
XX  
KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;  
KM ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20028534-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 19-NOV-2001; 2001WO-US43704.  
XX  
PR 20-NOV-2000; 2000US-252222P.  
PR 06-FEB-2001; 2001US-267011P.  
PR 28-MAR-2001; 2001US-279670P.  
PR 10-JUL-2001; 2001US-304037P.  
XX  
PA (CORI-) CORIXA CORP.  
PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;  
XX  
DR WPI; 2002-608400/65.  
XX  
PT New isolated tumor colon polynucleotide and polypeptide, useful for the  
PT diagnosis, prevention and/or treatment of cancer, in particular colon  
PT cancer -  
XX  
PS Claim 1; SEQ ID NO 578; 266bp + Sequence Listing; English.  
XX  
CC The invention relates to a human colon tumour expressed polynucleotide  
CC (I) encoding a polypeptide (II), ABP67991-ABP67996 comprising: (i) any of  
CC 2600 fully defined nucleotide sequences (ABV8669-ABV9289); (ii)  
CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)  
CC sequences that hybridize to (i), under moderately stringent conditions;  
CC (v) sequences having at least 75% or 90% identity to (i); or (vi)  
CC degenerate variants of (i). The compositions and methods of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC cancer, particularly colon cancer. (i) can be used in gene therapy and  
CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 486 BP; 192 A; 55 C; 64 G; 175 T; 0 other;

Query Match 68.1%; Score 467.8; DB 24; Length 486;  
Best Local Similarity 98.9%; Pred. No. 2.4e-71;  
Matches 469; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATTTTAATTTCTATTAACATCTTCTCAAGCATTAATTTATCCATATCTCACTGAA 60  
PS ATTTTAATTTCTATTAATATCTTCTCAAGCATTAATTTATCCATATCTCACTGAA 72  
Db 13 ATTTTAATTTCTATTAATATCTTCTCAAGCATTAATTTATCCATATCTCACTGAA 72

Qy 61 TTTTAAANAATAACATTAAGTATTTAGAAAACTAGAAAAAGATNAATGAGATAATTAA 120  
Db 73 TTTTAAAGAAATAACATTAAGTATTTAGAAAACTAGAAAAAGATNAATGAGATAATTAA 132  
Qy 121 ACTTACATGAAAAAGAAAAATTAATACAAAGAGCTGAGAACTGTAATTAATGAATGAGA 180  
Db 133 ACTTACATGAAAAAGAAAAATTAATACAAAGAGCTGAGAACTGTAATTAATGAATGAGA 192  
Qy 181 TTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGTCATTTATCTTAATGATGG 240  
Db 193 TTATTAATTTGAAAACTGATCTGAAAGCAAACTTTATTTGTCATTTATCTTAATGATGG 252  
Qy 241 TGTTTTATGACTAATACAGTATTTTCAATAGAAACCCAGTAAATAATTTTAT 300  
Db 253 TGTTTTATGACTAATACAGTATTTTCAATAGAAACCCAGTAAATAATTTTAT 312  
Qy 301 TTTTAAATAAAGCCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTGATTGGAGAGA 360  
Db 313 TTTTAAATAAAGCCTGTGTTCAAGCTCTGATCATATTTCTTTTATTTGATTGGAGAGA 372  
Qy 361 AAATACGTGTTTCTGATAGCATGAAATGCAAAATTTTATGATTTTAATCTCATAATTTT 420  
Db 373 AAATACGTGTTTCTGATAGCATGAAATGCAAAATTTTATGATTTTAATCTCATAATTTT 432  
Qy 421 AAGAACTATGAGAAATTTGATTATGACATGAACTGACAACTAATTAATCTGG 474  
Db 433 AAGAACTATGAGAAATTTGATTATGACATGAACTGACAACTAATTAATCTGG 486

RESULT 5  
AAX40097  
ID AAX40097 standard; DNA, 843 BP.  
XX  
AC AAX40097;  
XX  
DT 02-JUL-1999 (first entry)  
XX  
DE Gastric cancer associated gene.  
XX  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9904265-A2.  
XX  
PD 28-JAN-1999.  
XX  
PF 15-JUL-1998; 98WO-US14679.  
XX  
PR 22-JUN-1998; 98US-0102322.  
PR 17-JUL-1997; 97US-0896164.  
PR 10-OCT-1997; 97US-0061599.  
PR 10-OCT-1997; 97US-0061765.  
PR 10-OCT-1997; 97US-0948705.  
PR 11-OCT-1997; 97GB-0021697.  
XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;  
PI Tureci O;  
XX  
DR WPI; 1999-132448/11.  
XX  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
XX  
PS Claim 67; Page 696; 787bp; English.

CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.

XX Sequence 843 BP, 237 A, 184 C, 89 G, 215 T, 118 other;

Query Match 59.9%; Score 411.8; DB 20; Length 843;

Best Local Similarity 71.8%; Pred. No. 8.8e-62;

Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;

QY 1 ATTTAATTTCTATTAACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 60  
 DB 15 ATTTAATTTCTATTAACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 74  
 QY 61 TTTTAAATAAATACATTAAGTAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 120  
 DB 75 TTTTAAATAAATACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 134  
 QY 121 ACTTACATGAAGAAAGAAATTTATTAACAAAGACTGAGAACGTTATTAATGAATGAGA 180  
 DB 135 CTTTCTGAGAAAGAAAGAAATTTATTAACAAAGACTGAGAACGTTATTAATGAATGAGA 194  
 QY 181 TTATTAATTTGAAAACGTCATTTGAAAAGCAATTTATTTGTTCAATTAATTTGATGG 240  
 DB 195 TTATTAATTTGAAAACGTCATTTGAAAAGCAATTTATTTGTTCAATTAATTTGATGG 254  
 QY 241 TGTTTTATGACTATTAACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 300  
 DB 255 NNTTTTAAATAAATACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 314  
 QY 301 TTTTAAATAAATACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 360  
 DB 315 TTTTAAATAAATACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 374  
 QY 361 AAATACCTGTTCTGATAGCATGAAGAAATTTTATGATTTTATCTATCTCACTGAA 420  
 DB 375 AAATACCTGTTCTGATAGCATGAAGAAATTTTATGATTTTATCTATCTCACTGAA 434  
 QY 421 AAGAACTTATGAGAAATTTATGATAGCATGAAGAAATTTTATGATTTTATCTATCTG 480  
 DB 435 AAGAACTTATGAGAAATTTATGATAGCATGAAGAAATTTTATGATTTTATCTATCTG 493  
 QY 481 GTTGGCATTTGTTCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 540  
 DB 494 GTTGGCATTTGTTCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 553  
 QY 541 NGAATTAACCTTAAATATATCTTTGTTAGCCAAAC--AAACCTTTTGTGTTACATAGTCT 598  
 DB 554 AAATTAACCTTAAATATATCTTTGTTAGCCAAACCTTTTGTGTTACATAGTCT 613  
 QY 599 TTGATTTTATCTGTTCTTAAATTTATTTGTTGAAAGCTCCATTTTCCCGAGACATTAATTA 658  
 DB 614 TGGGATTTAAGGGGCTCCCAATTTATTCGAAACCAATTTTCCCGAAACCATTAATTA 673  
 QY 659 CCTATTAATTTGTTGTTATGACAGTTGTT 687  
 DB 674 CAT-TTTTACTTTGGTTAAGGCACAGTTGTT 701

RESULT 6  
 AAT19975/c  
 ID AAT19975 standard; cDNA to mRNA; 351 BP.

XX AAT19975;  
 AC 17-JUL-1996 (first entry)  
 DT 17-JUL-1996 (first entry)  
 XX Human gene signature HUMG01107.

XX Gene signature: messenger RNA; relative abundance; frequency;  
 KM human; cloning; mapping; non-biased library; diagnosis; detection;  
 KM cell typing; abnormal cell function; sb.

XX Homo sapiens.

XX W09514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94MO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATS/) MATSUBARA K.  
 PA (OKUB/) OKUBO K.

XX Matsubara K, Okubo K;  
 PI WPI: 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues

PS Claim 1, Page 525; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in AAT19901-726837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.

XX Sequence 351 BP, 117 A, 44 C, 44 G, 141 T, 5 other;

Query Match 43.8%; Score 300.8; DB 16; Length 351;

Best Local Similarity 97.0%; Pred. No. 7.8e-43;

Matches 324; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY 1 ATTTAATTTCTATTAACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 59  
 DB 334 ATTTAATTTCTATTAACATTTCTCTCAAGCATTTATTTATCTATCTCACTGAA 275  
 QY 60 ATTTTAAATAAATACATTTAGTATTAAGAAAACCTGAGAAAAGTAAATGAGATTAATTA 119  
 DB 274 ATTTTAAATAAATACATTTAGTATTAAGAAAACCTGAGAAAAGTAAATGAGATTAATTA 215  
 QY 120 AACTTACATGAAGAAAGAAATTTATTAACAAAGACTGAGAAAGTAAATGAGATTAATTA 179  
 DB 214 AACTTACATGAAGAAAGAAATTTATTAACAAAGACTGAGAAAGTAAATGAGATTAATTA 155  
 QY 180 ATTTAATTTGAAAACGTCATTTGAAAAGCAATTTATTTGTTCAATTAATTTATGATG 239  
 DB 154 ATTTAATTTGAAAACGTCATTTGAAAAGCAATTTATTTGTTCAATTAATTTATGATG 95

QY 240 GTGTTTATGACATAATACACTGATTTTTCATTAAGAAACCATGTTAAATATTT-TT 298  
 DB 94 GTGTTTATGACATAATACACTGATTTTTCATTAAGAAACCATGTTAAATATTTN 35  
 QY 299 ATTTTAAATAAGCCTGTGTTCAAGCTCTGATC 332  
 DB 34 NTTTTAAATAAGCCTGTGTTCAAGCTCTGATC 1

RESULT 7  
 AAH57293/c  
 ID AAH57293 standard; cDNA; 262 BP.  
 AC AAH57293;  
 XX  
 XX 10-SEP-2001 (first entry)  
 DT  
 DE Human liver specific cDNA sequence SEQ ID NO:133.  
 XX  
 XX Human; tissue specific; diagnosis; brain; heart; skeletal muscle;  
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;  
 KW metabolic disease; developmental disease; cytosstatic; immunomodulatory;  
 KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
 XX  
 OS Homo sapiens.  
 XX  
 XX MO200132927-A2.  
 PN  
 XX 10-MAY-2001.  
 PD  
 PF 02-NOV-2000; 2000WO-US30396.  
 XX  
 XX 04-NOV-1999; 99US-0163508.  
 PR  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX Sornasse T, Seilhamer JJ, Watson GA;  
 PI  
 XX WPI, 2001-291057/30.  
 DR  
 XX  
 XX New cell and tissue specific polynucleotides useful for diagnosis,  
 PT prognosis or monitoring of treatments for disorders where the gene is  
 associated with a cancer, immunopathology or neuropathology -  
 XX  
 XX  
 PS Claim 1; Page 117; 327pp; English.  
 CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
 CC sequences (I). (I) can have cytosstatic, immunomodulatory and  
 CC neuroprotective activities, and can be used in gene therapy. (I) and  
 CC proteins (II) encoded by then are used in high throughput screening  
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
 CC mimetics, peptides, proteins, agonists, antagonists, antibodies or  
 CC their fragments, immunoglobulins, inhibitors, drug compounds and  
 CC pharmaceutical agents. Expression of (I) in a sample indicates the  
 CC differentiation of embryonic stem cells into a tissue selected from  
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
 CC tissues. (I) and (II) are used to produce an expression profile that  
 CC defines a metabolic or developmental process, treatment, condition,  
 CC disease or disorder. The gene profile can be used for diagnosis,  
 CC prognosis or monitoring of treatments and for investigating a  
 CC predisposition to a disorder where the gene is associated with a  
 CC cancer, immunopathology or neuropathology.  
 CC  
 XX  
 SQ Sequence 262 BP; 103 A; 38 C; 34 G; 87 T; 0 other;

Query Match 38.0%; Score 261; DB 22; Length 262;  
 Best Local Similarity 99.6%; Pred. No. 4.8e-36;  
 Matches 261; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 243 TTTTATGACTAATACACGATTTTTCATTAAGAAACCATGTTAAATATTTTATTT 302  
 DB 262 TTTTATGACTAATACACGATTTTTCATTAAGAAACCATGTTAAATATTTTATTT 203

QY 303 TAAATAAGCGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTGATTTGGAGAAA 362  
 DB 202 TAAATAAGCGCTGTGTTCAAGCTCTGATCATATTTCTTTATTTGATTTGGAGAAA 143  
 QY 363 ATACTGTTTGTGATAGCATGAAATGCAAAATTTTGAATTTTATCTCNCATATTTTAA 422  
 DB 142 ATACTGTTTGTGATAGCATGAAATGCAAAATTTTGAATTTTATCTCNCATATTTTAA 83  
 QY 423 GAACATTTGAGAAATTTGATTAATGACATGAGTCACACACTAATTTACTGCGACGCTGT 482  
 DB 82 GAACATTTGAGAAATTTGATTAATGACATGAGTCACACACTAATTTACTGCGACGCTGT 23  
 QY 483 TGGCATTTGTGTTTCTTACTTAG 504  
 DB 22 TGGCATTTGTGTTTCTTACTTAG 1

RESULT 8  
 ABX83093/c  
 ID ABX83093 standard; cDNA; 173 BP.  
 AC ABX83093;  
 XX  
 XX 24-APR-2003 (first entry)  
 DT  
 DE Corn ear-derived polynucleotide (cpd) #1553.  
 XX  
 XX Corn ear-derived polynucleotide; cpd; cDNA library; SATMON022;  
 KW SATMON023; structural gene; functional gene; regulatory gene;  
 KW corn ear-specific profile; gene transcription; gene expression;  
 KW hybrid plant; desirable trait expression; plant breeding program;  
 KW inheritance; desirable characteristic; growth; development;  
 KW disease resistance; environmental adaptability; quality; yield;  
 KW multigene trait; plant; gene; ss.  
 XX  
 XX Zea mays.  
 OS  
 XX US6476212-B1.  
 PN  
 XX 05-NOV-2002.  
 PD  
 PF 14-MAY-1999; 99US-0313294.  
 XX  
 XX 26-MAY-1998; 98US-086722P.  
 PR  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX Laijudi RV, Ito LY, Sherman BK;  
 PI  
 XX WPI, 2003-208840/20.  
 DR  
 XX  
 XX Novel purified corn-ear derived polynucleotide useful as hybridization  
 PT probe for detecting polynucleotide in sample, and for identifying,  
 PT evaluating, and altering desired characteristics associated with  
 PT growth, development -  
 XX  
 XX  
 PS Examples; SEQ ID No 1553; 390pp; English.  
 CC  
 XX The present invention relates to the isolation of corn ear-derived  
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries  
 CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,  
 CC functional, and regulatory genes of corn ear. The polynucleotides  
 CC sequences are useful for detecting cpds in a sample, for producing  
 CC a corn ear-specific profile of gene transcription, for detecting  
 CC altered gene expression in inbred or hybrid plants, and for screening  
 CC several molecules for specific binding to the polynucleotide. The cpds  
 CC are useful to identify, isolate, or extend identical or related  
 CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic  
 CC acid amplification or hybridisation techniques to follow the  
 CC expression of desirable traits through plant breeding programs.  
 CC Preferably, the cpds are used to identify, evaluate, alter, or  
 CC follow the inheritance of desired characteristics associated with  
 CC growth and development, disease resistance, environmental adaptability,



CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 1512 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 1512 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC [seqdata.uspto.gov/sequence.html?docID=20020137139](http://seqdata.uspto.gov/sequence.html?docID=20020137139).

SQ Sequence 397 BP; 144 A; 75 C; 45 G; 133 T; 0 other;

Query Match 15.1%; Score 103.6; DB 25; Length 397;

Best Local Similarity 71.9%; Pred. No. 3.5e-09;

Matches 192; Conservative 0; Mismatches 65; Indels 10; Gaps 4;

OY 421 AAGAACTATTGAGAAATGATTAATGACATGAGTGCACAACTAATTACTGGCCAGCT 480  
 DB 14 AATTATCATCTAGAAATATATTAAAGTCTATTAAATGCACATTAATTACTGGTCAAGCT 73  
 OY 481 GTTGCAATGTGTTCTTCTTACTAGTCTCCCAAGAAAACCTTAACTGTAATCTTCAGC 540  
 DB 74 AATAATGATAT-TCGTCTTCTTCTTCTTCCCAAGAAAACCTTAACTGTAATCTTCAGC 125  
 OY 541 NGAAATACCTTAATAATTAATCTTTAGCCAAACAACTTTTGTGTACATAGTCTTT 600  
 DB 126 AGAATTAATCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 184  
 OY 601 GGATTTTACTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 660  
 DB 185 GTATTTTGTCTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 244  
 OY 661 TATTTTACTTTGTTATGACAGTGTGT 687  
 DB 245 TA-TTAACTTTGTATGACAGTGTGT 270

RESULT 11

ABI99762/c  
 ID ABI99762 standard; cDNA; 2152 BP.

XX AC AB199762;

DT 07-MAR-2002 (first entry)

XX DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:842.

XX KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 XX KM vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.

XX OS Mus musculus.

XX PN WO20018188-A2.

XX PD 22-NOV-2001.

XX PF 18-MAY-2001; 2001WO-JP04192.

XX PR 18-MAY-2000; 2000JP-0145977.

XX PA (UNIV-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX PF WPI, 2002-034733/04.

DR P-PSDB; ABB57301.

PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring

PT expression levels of particular genes defined in the specification or

PT by determining the expression profile of a gene group comprising these

PT genes -

PS Claim 2; Page 2087-2091; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (I) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (I). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (ABB57301 to ABB57374) or by determining the  
 CC the protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression levels of a gene group comprising these genes. The  
 CC expression levels or expression profiles produced by these genes are  
 CC used as an indicator when screening for ischaemic condition-improving  
 CC drugs or therapeutics for ischaemic diseases. ABB59913 and ABB59914  
 CC represent PCR primers for a mouse ischaemic condition related sequence,  
 CC which are used in the exemplification of the present invention.

XX SQ Sequence 2152 BP; 586 A; 455 C; 576 G; 535 T; 0 other;

Query Match 12.5%; Score 85.8; DB 24; Length 2152;

Best Local Similarity 65.2%; Pred. No. 3.8e-06;

Matches 172; Conservative 0; Mismatches 88; Indels 4; Gaps 3;

OY 421 AAGAACTATTGAGAAATGATTAATGACATGAGTGCACAACTAATTACTGGCCAGCT 480  
 DB 2130 AACTATTAGTGAAGAACTGATGACGATTAATGATGAGGACTAATCTGTGCTCACT 2011  
 OY 481 GTTGCAATGTGTTCTTCTTACTAGTCTCCCAAGAAAACCTTAACTGTAATCTTCAGC 540  
 DB 2070 -ATGAATCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2012  
 OY 541 NGAAATACCTTAATAATTAATCTTTAGCCAAACAACTTTTGTGTACATAGTCTTT 600  
 DB 2011 ACACTGATCTCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1954  
 OY 601 GGATTTTACTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 659  
 DB 1953 GATTTTACTAGTTTGAATTTTATTTCTGAAATCTAAGTTTAACCCCAACCATATTGAC 1854  
 OY 660 CTATTTAATCTTTGTTATGACAGT 683  
 DB 1993 ATAGTCAGCTTCATCAAGCAGT 1870

RESULT 12

ABZ10246  
 ID ABZ10246 standard; DNA; 8056 BP.

XX AC ABZ10246;

DT 16-JAN-2003 (first entry)

XX DE Haematopoietic cell proliferation disorder related DNA sequence #386.

XX KW Human; haematopoietic cell proliferation disorder; cytostatic;  
 XX KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 XX KM cytosine methylation state; gene; ds.

XX OS Homo sapiens.

XX PN WO20027272-A2.

XX PD 03-OCT-2002.

XX PF 26-MAR-2002; 2002WO-EP03401.

XX

```

PR 26-MAR-2001; 2001US-278333P.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Berlin K, Braun A, Distler J, Guefig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu B;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
PI Petet C, Schwöpe I, Ziebarth H;
XX
XX WPI; 2003-018942/01.
XX
XX Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent
PT that distinguishes between methylated and non-methylated CpG
PT dinucleotides -
XX
XX Claim 28; SEQ ID 386; 117bp; English.
XX
XX The present invention describes a method for detecting and
CC differentiating between hematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. AB209861 to AB21118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used for
CC differentiating between healthy hematopoietic cells and proliferative
CC disorder hematopoietic cells; for differentiating between acute
CC lymphocytic leukemia and acute myelogenous leukemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of hematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of hematopoietic cell proliferation disorder related
CC DNA sequences. The nucleotide sequences from the present invention can
CC also be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC hematopoietic cell proliferative disorders. The present method enables
CC a highly specific classification of hematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients.
XX
XX Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
SQ
Query Match 10.4%; Score 71.6; DB 25; Length 8056;
Best Local Similarity 50.3%; Pred. No. 0.00098;
Matches 224; Conservative 0; Mismatches 217; Indels 4; Gaps 2;
XX
OY 2 TTTAATTTCTAATTAACATTCCTCGAAGCATTATTTATCCTATATCTCACTGAT 61
DB 2056 TTAATAAATTTATTTAATAATTTAATTTAATTAATTAATTAATTAATTTAT 2115
OY 62 TTTAANAATAACATTAAGTATTAAGAAAACTAGAAAAAGATNAATGACATAATTA 121
DB 2116 ATAATTTTAATTAATTTTATTTAATTAATAATAATAATAATAATAATAATTA 2175
OY 122 CTTACATGAAGAAAGAAATTAAT--ACAAAGAGCTGGAACGTTAATAATTGAAATG 179
DB 2176 TTTTAATTAATAAATTTAATAAATAAATAAATAAATAAATAAATAAATTAATTT 2235
OY 180 ATTAATTAATTTGAAAGCTCATCTGAAAGCAACTTATTTGTCATTTCTTAATGATG 239
DB 2236 ATAAATAATTTATTTTAAATTTAATAATAATAATAATAATAATAATAATAATTT 2255
OY 240 GTGTTTATGACATAACATGATTTTTCATTAAGAAACCAGTGTAAATAATTTTAA 299
DB 2296 TTAATTTTAATAATTAATTAATTTT--GTAATTAATAATTAATTTTAAATTTTAA 2353
OY 300 TTTTAATAATAAGCTGCTGCTCAAGCTGATCATATTTCTTTATTTGATTTGGAG 359
DB 2354 ATATTAATTTTATTTTATTAATAATTAATTAATTTTATTTATTTTGAATAA 2413
OY 360 AAAATACGTGTTTCGATGATGAAATGCAAAATTTTATGATTTTATTCNCATATTT 419

```

```

DB 2414 AAAAAAATTAATTTATTTTAAATAATTAATTTTATTTTATTTATTAATTTT 2473
OY 420 TAAGAACTATTGAGAAATGATTA 444
DB 2474 TATTAATTTGTTTAAATAATTA 2498
XX
XX RESULT 13
XX AB210246/C
ID AB210246 standard; DNA; 8056 BP.
XX
XX AB210246;
AC
XX 16-JAN-2003 (first entry)
DT
XX
XX Haematopoietic cell proliferation disorder related DNA sequence #386.
DE
XX
XX Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukemia; acute myelogenous leukemia;
KW cytosine methylation state; gene; dg.
XX
XX Homo sapiens.
OS
XX
XX W0200277272-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 26-MAR-2002; 2002MO-EP03401.
PF
XX
XX 26-MAR-2001; 2001US-278333P.
PR
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Berlin K, Braun A, Distler J, Guefig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu B;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
PI Petet C, Schwöpe I, Ziebarth H;
XX
XX WPI; 2003-018942/01.
XX
XX Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent
PT that distinguishes between methylated and non-methylated CpG
PT dinucleotides -
XX
XX Claim 28; SEQ ID 386; 117bp; English.
XX
XX The present invention describes a method for detecting and
CC differentiating between hematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. AB209861 to AB21118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used for
CC differentiating between healthy hematopoietic cells and proliferative
CC disorder hematopoietic cells; for differentiating between acute
CC lymphocytic leukemia and acute myelogenous leukemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of hematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of hematopoietic cell proliferation disorder related
CC DNA sequences. The nucleotide sequences from the present invention can
CC also be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC hematopoietic cell proliferative disorders. The present method enables
CC a highly specific classification of hematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients.
XX
XX Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
SQ
Query Match 10.2%; Score 70; DB 25; Length 8056;

```

Best Local Similarity 46.9%; Pred. No. 0.0018;  
Matches 320; Conservative 0; Mismatches 353; Indels 10; Gaps 3;

```

QY 6 AATTTTCTATTAACACTTCTCTCAAGCATTTATTTATCCATATCTCAGTGAATTTTA 65
   |||||
DB 1019 AATTTTAAATATATTTTAAAAAAATTAATAATATTTTATTCATTTAAATTA 960
   |||||

QY 66 AAAAAATTAACATAGATTAAGAAAACTAGAAAAAGNNATGAGATTAATAACTTA 125
   |||||
DB 959 AATTAATAATTTTATTTTAAAAAAATTTATTAACAAAAACAAAAATTAATAAT 900
   |||||

QY 126 CATGAAAAAGAAAAATTAACAAAGACTGAGAAAGTATTAATGAA-ATGAGATTAT 184
   |||||
DB 899 AATTTTAAATATTCAAAAATTTTATACATTAATAAAAAATTAATAAACATTTT 840
   |||||

QY 185 AATTTGAAAACCTGCATCTGAAAGCAACTTTATTTGTCATATTTCTTAATGATGCTGT 244
   |||||
DB 839 AATTAATAATTTTAAATTAATCATTTATTAATAATTTTAAATATATATTAAT 780
   |||||

QY 245 TTATGACTAATACACGTATTTTCAATAAGAAACCCATTAATAATTTTATTTTA 304
   |||||
DB 779 TTTATTAATTTTATTAATTTTAAAAATTTATTAATTAATTAATTAATTTTAA 720
   |||||

QY 305 AAAATTAAGCTGTGTCAAGCTCTGATCATATTTCTTTATTTGATTGGAGAAAT 364
   |||||
DB 719 AATTAATTTTAAATTAATAAAATTTTATTAATTTTATTTTATTTTAAATA 660
   |||||

QY 365 ACTGTTCTGATGATGATGAATGCAAAATTTTGAATTTTAATCTCCTAATTTTAA 424
   |||||
DB 659 AATTAATTTTAAATTAATTTTAAAAATTTTAAAAATTTTAAATATTTTATTA 600
   |||||

QY 425 ACTATGGAATGATTAATGACATGAAGTGCACACATTAATCTGAGCAGCTGTG 484
   |||||
DB 599 AATTAATTTTAAATTAATAATTAATTTTAAATAAAACATTAATTAATTAATAATTT 540
   |||||

QY 485 GCATGTGCTTTCTTAATTAATTAATTTTCCCAAGAAACCTTAATGATCTTCAGCGAA 544
   |||||
DB 539 AATTTTATATATAA--TATTTAATATTAATAATTTTAAATTTTAAATTTT-----T 489
   |||||

QY 545 TAACCTTAATTAATTAATTTGTTAGCCAAACAAACTTTTGTTCATAGTTCTTTGAT 604
   |||||
DB 488 TTTATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 429
   |||||

QY 605 TTTACTGTCCATTTTATTTCTGAACCTTATTTTCCGAGCATAATTAACCTAT 664
   |||||
DB 428 AAAAAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTCA 369
   |||||

QY 665 TAACCTTGTATGACAGTTGTT 687
   |||||
DB 368 TTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 346
   |||||

```

## RESULT 14

ABL70409  
ID ABL70409 standard; DNA; 7676 BP.

AC ABL70409;

DT 01-JUL-2002 (first entry)

DE Chemically treated cell signalling DNA sequence#150.

KW Cell signalling; cytosine methylation; cell signalling disease;

KW cancer; tumour; cytosolic; ds.

OS Unidentified.

PN WO200202807-A2.

PD 10-JAN-2002.

PF 29-JUN-2001; 2001WO-EP07471.

XX

PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI, 2002-154758/20.

PT Nucleic acid, useful for diagnosis and therapy of diseases associated  
PT with cell signalling e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with cell signalling -

PS Claim 1; SEQ ID NO 299; 24bp-sequence listing; English.

CC The invention relates to a nucleic acid comprising a sequence of at least  
CC 18 bases of a segment of chemically pre-treated DNA of genes associated  
CC with cell signalling. The activity of the modified sequences of the  
CC invention may be described as cytosolic. The object of the invention is  
CC to provide the chemically modified DNA of genes associated with cell  
CC signalling, as well as oligonucleotides and/or PNA-oligomers for  
CC detecting cytosine methylations, as well as a method which is  
CC particularly suitable for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with cell signalling. The  
CC chemically modified DNA provided by the invention is useful for diagnosis  
CC and therapy of diseases such as solid tumours and cancer. The sequences  
CC given in records ABL70111-ABL70626 represent chemically pre-treated  
CC genomic DNA's of genes associated with cell signalling.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.

XX Sequence 7676 BP; 2396 A; 45 C; 1348 G; 3887 T; 0 other;

SQ Query Match 9.7%; Score 66.8; DB 24; Length 7676;

Best Local Similarity 48.5%; Pred. No. 0.0065;

Matches 210; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

```

QY 21 ATTCCTTCAAGCATTTATTTATCTCATATCTCAGTGAATTTTAAATAATACATT--- 77
   |||||
DB 3697 AATTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3756
   |||||

QY 78 AGTATTAAGAAAACTAGAAAAAGATTAATGCAATTAATTAATTAATTAATTAATTA 137
   |||||
DB 3757 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3816
   |||||

QY 138 AATTTAATTAAGAAAGCTGAGAAAGTTATTAATTAATTAATTAATTAATTAATTA 197
   |||||
DB 3817 AATGTATGTTAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 3876
   |||||

QY 198 CATCTGAAGCAACCTTAATTTGTCATTAATTTCTTAATTAATTAATTAATTAATTA 257
   |||||
DB 3877 AATTTAATTAATTAATTAATTTTATTTATTAATTAATTAATTTTATTTATTAAT 3936
   |||||

QY 258 ACGATTTTCAATTAAGAAAGCCATTAATTAATTAATTAATTAATTAATTAATTA 317
   |||||
DB 3937 GATTAATTTTATTTTAAAGTAAAGTTATTAATTAATTTTAAAGTAAATTAATTA 3996
   |||||

QY 318 GTTCAAGCTCTGATCAATTTCTTTATTTGATTTGGAAGAAATTAATTAATTAATTA 377
   |||||
DB 3997 TTTTGTGTGTGTTTATTTTATTTGAAATATTTTGTGTTGAAATTAATTAATTAAT 4056
   |||||

QY 378 GCATGAATGCAAAATTTTGAATTTTAAATCTCNCCTAATTTTAAAGAACTATTA 437
   |||||
DB 4057 AATTAATTAATTAATTAATTTTATTAATTTTAAATTAATTAATTAATTAATTA 4116
   |||||

QY 438 TGATTAATGACAT 450
   |||||
DB 4117 AGTAAATAGTAT 4129
   |||||

```

## RESULT 15

ABL34598





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 06:06:02 : Search time 259.057 Seconds  
(without alignments)  
8737.270 Million cell updates/sec

Title: US-09-835-992A-20

Perfect score: 687

Sequence: 1 atttatacttctcttaaac.....ctctgtatgcacagtcgtc 687

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues 4380138

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683	92.9	687	9	US-09-835-992A-20
2	638.4	92.9	1052	10	US-09-880-107-3296
3	611.8	89.1	714	9	US-09-835-992A-19
4	471	68.6	506	10	US-09-796-692-7228
5	471	68.6	506	14	US-10-040-862-7228
6	467.8	68.1	486	10	US-09-998-598-578
7	411.8	59.9	843	9	US-09-835-992A-22
8	161.8	23.6	510	14	US-10-102-524-161
9	161.8	23.6	510	14	US-10-102-524-188
10	161.8	23.6	510	14	US-10-102-524-645
11	103.6	15.1	397	10	US-09-960-352-5962
12	66.8	9.7	7676	12	US-10-240-485-151
13	65	9.5	113515	12	US-10-311-455-2147
14	63.4	9.60	960	14	US-10-198-846-6381
15	63	9.2	6106	12	US-10-311-455-1445
16	62	9.0	3673778	12	US-10-312-841-1

C 17	61.8	9.0	291	10	US-09-960-352-13694	Sequence 13694, A
C 18	60.4	8.8	6944	14	US-10-172-086-111	Sequence 111, App
C 19	60.2	8.8	6509	12	US-10-311-455-199	Sequence 199, App
C 20	60.2	8.8	6912	12	US-10-240-452-245	Sequence 245, App
C 21	59.8	8.7	17721	12	US-10-311-455-1702	Sequence 1702, App
C 22	59.6	8.7	10467	12	US-10-240-453-327	Sequence 327, App
C 23	59.4	8.6	3991	14	US-10-074-045-60	Sequence 60, App1
C 24	59.2	8.6	640681	10	US-09-790-988-1	Sequence 1, App1
C 25	59	8.6	3673778	12	US-10-312-841-2	Sequence 2, App1
C 26	58.8	8.6	449	10	US-09-960-352-11260	Sequence 11260, A
C 27	58.4	8.5	424	10	US-09-960-352-11218	Sequence 11218, A
C 28	58.2	8.5	5962	12	US-10-311-455-1260	Sequence 1260, App
C 29	58	8.4	18218	12	US-10-311-455-1922	Sequence 1922, App
C 30	57.8	8.4	9219	12	US-10-311-455-2400	Sequence 2400, App
C 31	57.8	8.4	3673778	12	US-10-312-841-1	Sequence 1, App1
C 32	57.6	8.4	9810	12	US-10-311-455-400	Sequence 400, App
C 33	57.6	8.4	3673778	12	US-10-312-841-2	Sequence 2, App1
C 34	57.2	8.3	640681	10	US-09-790-988-1	Sequence 1, App1
C 35	57	8.3	480	10	US-09-960-352-5301	Sequence 5301, App
C 36	57	8.3	13712	12	US-10-311-455-1504	Sequence 1504, App
C 37	56.8	8.3	11745	12	US-10-240-453-206	Sequence 206, App
C 38	56.8	8.3	18624	12	US-10-311-455-1675	Sequence 1675, App
C 39	56.2	8.2	416	10	US-09-960-352-4584	Sequence 4584, App
C 40	56.2	8.2	8979	12	US-10-311-455-758	Sequence 758, App
C 41	56.2	8.2	12138	12	US-10-311-455-1916	Sequence 1916, App
C 42	56.2	8.2	12138	12	US-10-240-453-210	Sequence 210, App
C 43	55.8	8.1	5276	12	US-10-311-455-124	Sequence 124, App
C 44	55.8	8.1	7049	12	US-10-311-455-130	Sequence 130, App
C 45	55.8	8.1	7049	12	US-10-240-452-6	Sequence 6, App1

## ALIGNMENTS

RESULT 1

US-09-835-992A-20

Sequence 20, Application US/09835992A

Patent No. US20020037541A1

GENERAL INFORMATION:

APPLICANT: Obata, Yuichi

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND

FILE REFERENCE: 10461/7112

CURRENT APPLICATION NUMBER: US/09/835, 992A

CURRENT FILING DATE: 2001-04-16

PRIOR FILING DATE: 1997-07-17

NUMBER OF SEQ ID NOS: 87

SOFTWARE: PatentIn version 3.0

SEQ ID NO 20

LENGTH: 687

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: Unsure

LOCATION: (67)..(67)

OTHER INFORMATION: n = a, c, g or t

NAME/KEY: Unsure

LOCATION: (105)..(105)

OTHER INFORMATION: n = a, c, g or t

NAME/KEY: Unsure

LOCATION: (412)..(412)

OTHER INFORMATION: n = a, c, g or t

NAME/KEY: Unsure

LOCATION: (541)..(541)

OTHER INFORMATION: n = a, c, g or t

US-09-835-992A-20

Query Match 99.4%; Score 683; DB 9; Length 687;

Best Local Similarity 100.0%; Pred. No. 5.8e-114;

Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATTTAATTTTCATTAACATCTTCTCAAGCATATATTATTCATATCTCATGAA 60

```

Db      1 ATTTTAATTTCTATTAACATCTCTCCAAAGCATTTATTCCTATATCTCACTGAA 60
Qy      61 TTTTAAANAATPAACATAGTATTAAGAAAAAAGTAAAAAGATNAATGACATATTTAA 120
Db      61 TTTTAAANAATPAACATAGTATTAAGAAAAAAGTAAAAAGATNAATGACATATTTAA 120
Qy      121 ACTTACATGAAAAAGAAAAATTTATPAACAAGACAGTAAAGCTTAAATTTGAATGGA 180
Db      121 ACTTACATGAAAAAGAAAAATTTATPAACAAGACAGTAAAGCTTAAATTTGAATGGA 180
Qy      181 TTATATTTGAAAACTGCATCTGAAAGCAAACTTTATTTGTTCAATTATTTCTTAATGATG 240
Db      181 TTATATTTGAAAACTGCATCTGAAAGCAAACTTTATTTGTTCAATTATTTCTTAATGATG 240
Qy      241 TGTTTTATGACTAATACATGATTTTTCATTAAGAAAAACCATGTTAAATAATTTTAT 300
Db      241 TGTTTTATGACTAATACATGATTTTTCATTAAGAAAAACCATGTTAAATAATTTTAT 300
Qy      301 TTTTAAATAAGCCCTGCTTGAAGCTGCATCATATTTCTTTATTTGATTTGGAGAA 360
Db      301 TTTTAAATAAGCCCTGCTTGAAGCTGCATCATATTTCTTTATTTGATTTGGAGAA 360
Qy      361 AAATACGTTCTGATAGCATGAATAAGCAAAATTTTATTTTATCTGCTAATTTT 420
Db      361 AAATACGTTCTGATAGCATGAATAAGCAAAATTTTATTTTATCTGCTAATTTT 420
Qy      421 AAGAATCTATGAGAAATGATTAATGACATGAAGTGCACAACTAATTTACTGCGCAGCT 480
Db      421 AAGAATCTATGAGAAATGATTAATGACATGAAGTGCACAACTAATTTACTGCGCAGCT 480
Qy      481 GTTGCACTGTGTTCTTACTAGTCTCCCAAGAAAACTTAAATCTGAATCTTGAGC 540
Db      481 GTTGCACTGTGTTCTTACTAGTCTCCCAAGAAAACTTAAATCTGAATCTTGAGC 540
Qy      541 NGATAAACCTTAATATATCTTTGTAGCCAAACAACCTTTTGTTTACATAGTCTTT 600
Db      541 NGATAAACCTTAATATATCTTTGTAGCCAAACAACCTTTTGTTTACATAGTCTTT 600
Qy      601 GGATTTTACTGTTCTTAATTTTATTCGAAAATCCCATTTTCCCAAGCATTAATACC 660
Db      601 GGATTTTACTGTTCTTAATTTTATTCGAAAATCCCATTTTCCCAAGCATTAATACC 660
Qy      661 TATTTAATCTTGTATGACAGCTTGT 687
Db      661 TATTTAATCTTGTATGACAGCTTGT 687

```

## RESULT 2

```

US-09-880-107-3296/c
; Sequence 3296, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-NO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3296
; LENGTH: 1052
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U11313

```

## US-09-880-107-3296

Query Match 92.9%; Score 638.4; DB 10; Length 1052;  
 Best Local Similarity 98.1%; Pred. No. 7,4e-106;  
 Matches 675; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

```

Qy      1 ATTTTAATTTCTATTAACATCTCTCCAAAGCATTTATTCCTATATCTCACTGAA 60
Db      1014 ATTTTAATTTCTATTAACATCTCTCCAAAGCATTTATTCCTATATCTCACTGAA 955
Qy      61 TTTTAAANAATPAACATAGTATTAAGAAAAAAGTAAAAAGATNAATGACATATTTAA 120
Db      954 TTTTAAANAATPAACATAGTATTAAGAAAAAAGTAAAAAGATNAATGACATATTTAA 895
Qy      121 ACTTACATGAAAAAGAAAAATTTATPAACAAGACAGTAAAGCTTAAATTTGAATGGA 180
Db      894 ACTTACATGAAAAAGAAAAATTTATPAACAAGACAGTAAAGCTTAAATTTGAATGGA 835
Qy      181 TTATATTTGAAAACTGCATCTGAAAGCAAACTTTATTTGTTCAATTATTTCTTAATGATG 240
Db      834 TTATATTTGAAAACTGCATCTGAAAGCAAACTTTATTTGTTCAATTATTTCTTAATGATG 775
Qy      241 TGTTTTATGACTAATACATGATTTTTCATTAAGAAAAACCATGTTAAATAATTTTAT 300
Db      774 TGTTTTATGACTAATACATGATTTTTCATTAAGAAAAACCATGTTAAATAATTTTAT 715
Qy      301 TTTTAAATAAGCCCTGCTTGAAGCTGCATCATATTTCTTTATTTGATTTGGAGAA 360
Db      714 TTTTAAATAAGCCCTGCTTGAAGCTGCATCATATTTCTTTATTTGATTTGGAGAA 655
Qy      361 AAATACGTTCTGATAGCATGAATAAGCAAAATTTTATTTTATCTGCTAATTTT 420
Db      654 AAATACGTTCTGATAGCATGAATAAGCAAAATTTTATTTTATCTGCTAATTTT 595
Qy      421 AAGAATCTATGAGAAATGATTAATGACATGAAGTGCACAACTAATTTACTGCGCAGCT 480
Db      594 AAGAATCTATGAGAAATGATTAATGACATGAAGTGCACAACTAATTTACTGCGCAGCT 535
Qy      481 GTTGCACTGTGTTCTTACTAGTCTCCCAAGAAAACTTAAATCTGAATCTTGAGC 540
Db      534 GTTGCACTGTGTTCTTACTAGTCTCCCAAGAAAACTTAAATCTGAATCTTGAGC 475
Qy      541 NGATAAACCTTAATATATCTTTGTAGCCAAACAACCTTTTGTTTACATAGTCTTT 598
Db      474 NGATAAACCTTAATATATCTTTGTAGCCAAACAACCTTTTGTTTACATAGTCTTT 415
Qy      599 TTGATTTTACTGTTCTTAATTTTATTCGAAAATCCCATTTTCCCAAGCATTAATACC 658
Db      414 TTGATTTTACTGTTCTTAATTTTATTCGAAAATCCCATTTTCCCAAGCATTAATACC 355
Qy      659 CTTATTTAATCTTGTATGACAGCTTGT 686
Db      354 CATA-TTAACCTTGTATGACAGCTTGT 328

```

## RESULT 3

```

US-09-835-992A-19
; Sequence 19, Application US/09835992A
; Patent No. US20020037541A1
; GENERAL INFORMATION:
; APPLICANT: Obata, Yuichi
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND
; FILE REFERENCE: 10461/7112
; CURRENT APPLICATION NUMBER: US/09/835,992A
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: US 08/896,164
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 19
; LENGTH: 714
; TYPE: DNA

```

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Unsure  
LOCATION: (243) .. (243)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (373) .. (373)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (437) .. (437)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (702) .. (702)  
OTHER INFORMATION: n = a, c, g or t  
US-09-835-992a-19

Query Match 89.1%; Score 611.8; DB 9; Length 714;  
Best Local Similarity 97.1%; Pred. No. 4,1e-101;  
Matches 671; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

1 ATTTAAATTTTCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 60  
15 ATTTAAATTTTCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 74  
61 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 120  
75 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 134  
121 ACTTACATGAAAAGAAAATTTATTAACAAAGCATGAAAGTAAATGAGATTA 180  
135 ACTTACATGAAAAGAAAATTTATTAACAAAGCATGAAAGTAAATGAGATTA 194  
181 TTTTAAATTTGAAAAGTGAATGAAAGTAAATTTATTTGATTAATGATG 240  
195 TTTTAAATTTGAAAAGTGAATGAAAGTAAATTTATTTGATTAATGATG 254  
241 TGTATTAATGCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 300  
255 TGTATTAATGCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 314  
301 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 360  
315 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 374  
361 AAATACGTTTCTGATAGCATGAAAGTAAATTTATTTGATTAATGATG 420  
375 AAATACGTTTCTGATAGCATGAAAGTAAATTTATTTGATTAATGATG 434  
421 AAATACGTTTCTGATAGCATGAAAGTAAATTTATTTGATTAATGATG 480  
435 AAATACGTTTCTGATAGCATGAAAGTAAATTTATTTGATTAATGATG 494  
481 GTTGGCATTTGTTTCTTACTTATTTCTCCAA- GGAAGCTCTTAATCTTCA 539  
495 GTTGGCATTTGTTTCTTACTTATTTCTCCAAAGGAAAAGTCTTAATTTCA 554  
540 GGAAGCTCTTAAATTAATTTGTTTCTTACTTATTTCTCCAAAGGAAAAGT 597  
555 GGAAGCTCTTAAATTAATTTGTTTCTTACTTATTTCTCCAAAGGAAAAGT 614  
598 TTT- GGAATTTTACTGTTCTTAATTTATTTGTTTCTTACTTATTTCTCCAA 656  
615 TTTGGGATTTTACTGTTCTTAATTTATTTGTTTCTTACTTATTTCTCCAA 674  
657 ACCCTATTTTACTGTTCTTAATTTATTTGTTTCTTACTTATTTCTCCAA 687  
675 ACCATA- TTAATTTTCTGTTTCTTACTTATTTCTCCAAAGGAAAAGT 704

RESULT 4  
US-09-796-692-7228  
Sequence 7228, Application US/09796692  
Publication No. US20020198362A1

GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Maimon, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIORITY APPLICATION NUMBER: 60/186,126  
PRIORITY FILING DATE: 2000-03-01  
PRIORITY APPLICATION NUMBER: 60/190,479  
PRIORITY FILING DATE: 2000-03-17  
PRIORITY APPLICATION NUMBER: 60/200,545  
PRIORITY FILING DATE: 2000-04-27  
PRIORITY APPLICATION NUMBER: 60/200,303  
PRIORITY FILING DATE: 2000-04-28  
PRIORITY APPLICATION NUMBER: 60/200,779  
PRIORITY FILING DATE: 2000-04-28  
PRIORITY APPLICATION NUMBER: 60/200,999  
PRIORITY FILING DATE: 2000-05-01  
PRIORITY APPLICATION NUMBER: 60/202,084  
PRIORITY FILING DATE: 2000-05-04  
PRIORITY APPLICATION NUMBER: 60/206,201  
PRIORITY FILING DATE: 2000-05-22  
PRIORITY APPLICATION NUMBER: 60/218,950  
PRIORITY FILING DATE: 2000-07-14  
PRIORITY APPLICATION NUMBER: 60/222,903  
PRIORITY FILING DATE: 2000-08-03  
PRIORITY APPLICATION NUMBER: 60/223,416  
PRIORITY FILING DATE: 2000-08-04  
PRIORITY APPLICATION NUMBER: 60/223,378  
PRIORITY FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 7228  
LENGTH: 506  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-7228

Query Match 68.6%; Score 471; DB 10; Length 506;  
Best Local Similarity 99.4%; Pred. No. 8.8e-76;  
Matches 471; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATTTAAATTTTCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 60  
33 ATTTAAATTTTCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 92  
61 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 120  
93 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 152  
121 ACTTACATGAAAAGAAAATTTATTAACAAAGCATGAAAGTAAATGAGATTA 180  
153 ACTTACATGAAAAGAAAATTTATTAACAAAGCATGAAAGTAAATGAGATTA 212  
181 TTTTAAATTTGAAAAGTGAATGAAAGTAAATTTATTTGATTAATGATG 240  
213 TTTTAAATTTGAAAAGTGAATGAAAGTAAATTTATTTGATTAATGATG 272  
241 TGTATTAATGCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 300  
273 TGTATTAATGCTATTAACATTTCTCAAGCATTTATTCCTATATTCACAGAA 332  
301 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 360  
333 TTTTAAATAATTAACATTTAGTATTAAGAAAAGTAAATGAGATTAATTA 392  
361 AAATACGTTTCTGATAGCATGAAAGTAAATTTATTTGATTAATGATG 420  
393 AAATACGTTTCTGATAGCATGAAAGTAAATTTATTTGATTAATGATG 452

QY 421 AAGACTATTGAGAAATGATTAAATGACATGAGAGCAACACCTAATTACTGG 474  
DB 453 AAGAACTATTGAGAAATGATTAAATGACATGAGAGCAACACCTAATTACTGG 506

RESULT 5  
US-10-040-862-7228  
; Sequence 7228, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; FILE REFERENCE: 014058-013520US  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; PRIOR FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7228  
; LENGTH: 506  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-040-862-7228

Query Match 68.6%; Score 471; DB 14; Length 506;  
Best Local Similarity 99.4%; Pred. No. 8.8e-76;  
Matches 471; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATTTAATTTTCATTAACATCTCTCTCAAGCATTATTTTATCTATATCTCACTGAA 60  
DB 33 ATTTAATTTTCATTAACATCTCTCTCAAGCATTATTTTATCTATATCTCACTGAA 92  
QY 61 TTTTANAATAATACATTAGTTTGAAGAAAGTAAATGAGATTAATTA 120  
DB 93 TTTTANAATAATACATTAGTTTGAAGAAAGTAAATGAGATTAATTA 152  
QY 121 ACTTACATGAAAAAGAAATTAACAAGAGTGAAGCGTATATTAATGAAATAGA 180  
DB 153 ACTTACATGAAAAAGAAATTAACAAGAGTGAAGCGTATATTAATGAAATAGA 212  
QY 181 TTAATATTTGAAAACTGCATCTGAAAGCAAACTTATTTGTTCAATTTCTTAATGATG 240  
DB 213 TTAATATTTGAAAACTGCATCTGAAAGCAAACTTATTTGTTCAATTTCTTAATGATG 272

QY 241 TGTTTATGACTAATACTAGTATTTTCAATAGAGAAACCCAGTTAAATATTTTAT 300  
DB 273 TGTTTATGACTAATACTAGTATTTTCAATAGAGAAACCCAGTTAAATATTTTAT 332  
QY 301 TTTTAAAAATAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTGAATTTGGAGA 360  
DB 333 TTTTAAAAATAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTGAATTTGGAGA 392  
QY 361 AAATCTGTTTCTGATAGCATGAAATGCAAAATTTTGAATTTTGAATTTCTCNCATTTT 420  
DB 393 AAATCTGTTTCTGATAGCATGAAATGCAAAATTTTGAATTTTGAATTTCTCNCATTTT 452

QY 421 AAGACTATTGAGAAATGATTAAATGACATGAGAGCAACACCTAATTACTGG 474  
DB 453 AAGAACTATTGAGAAATGATTAAATGACATGAGAGCAACACCTAATTACTGG 506

RESULT 6  
US-09-998-598-578  
; Sequence 578, Application US/09998598  
; Patent No. US20020150922A1  
; GENERAL INFORMATION:  
; APPLICANT: Stolk, John A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Chenault, Ruth A.  
; APPLICANT: Meagher, Madelein Joy  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.561  
; CURRENT APPLICATION NUMBER: US/09/998,598  
; PRIOR FILING DATE: 2001-11-16  
; NUMBER OF SEQ ID NOS: 2606  
; SOFTWARE: Corixa Invention Disclosure Database  
; SEQ ID NO 578  
; LENGTH: 486  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-998-598-578

Query Match 68.1%; Score 467.8; DB 10; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.3e-75;  
Matches 469; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATTTAATTTTCATTAACATCTCTCTCAAGCATTATTTTATCTATATCTCACTGAA 60  
DB 13 ATTTAATTTTCATTAACATCTCTCTCAAGCATTATTTTATCTATATCTCACTGAA 72  
QY 61 TTTTANAATAATACATTAGTTTGAAGAAAGTAAATGAGATTAATTA 120  
DB 73 TTTTANAATAATACATTAGTTTGAAGAAAGTAAATGAGATTAATTA 132  
QY 121 ACTTACATGAAAAAGAAATTAACAAGAGTGAAGCGTATATTAATGAAATAGA 180  
DB 133 ACTTACATGAAAAAGAAATTAACAAGAGTGAAGCGTATATTAATGAAATAGA 192  
QY 181 TTAATATTTGAAAACTGCATCTGAAAGCAAACTTATTTGTTCAATTTCTTAATGATG 240  
DB 193 TTAATATTTGAAAACTGCATCTGAAAGCAAACTTATTTGTTCAATTTCTTAATGATG 252  
QY 241 TGTTTATGACTAATACTAGTATTTTCAATAGAGAAACCCAGTTAAATATTTTAT 300  
DB 253 TGTTTATGACTAATACTAGTATTTTCAATAGAGAAACCCAGTTAAATATTTTAT 312  
QY 301 TTTTAAAAATAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTGAATTTGGAGA 360  
DB 313 TTTTAAAAATAGCCTGTGTTCAAGCTCTGATCATATTTCTTTATTTTGAATTTGGAGA 372  
QY 361 AAATCTGTTTCTGATAGCATGAAATGCAAAATTTTGAATTTTGAATTTCTCNCATTTT 420  
DB 373 AAATCTGTTTCTGATAGCATGAAATGCAAAATTTTGAATTTTGAATTTCTCNCATTTT 432  
QY 421 AAGACTATTGAGAAATGATTAAATGACATGAGAGCAACACCTAATTACTGG 474



```
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (262)..(263)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (270)..(270)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (274)..(274)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (287)..(287)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (292)..(292)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (296)..(298)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (305)..(305)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (307)..(307)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (312)..(312)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (329)..(329)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (332)..(332)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (342)..(342)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (348)..(348)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (357)..(357)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (379)..(379)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (381)..(381)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (387)..(388)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (394)..(395)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (399)..(400)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (406)..(406)

Query Match 59.9%; Score 411.8; DB 9; Length 843;
Best Local Similarity 71.8%; Pred. No. 4.9e-65;
Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;
```

```
QY 121 ACTTACATGAAAAAGAAATTTATACAAAGACTGGAAGCTTATTAATGAAATGCA 180
DB 135 CTTTCTGAAAANGGAATTTTACAAAGACNGGAANCTTTATTTGAAATGAA 194
QY 181 TTATTAATTTGAAAACGATCTGAAAGCAACTTATTTGTTCAATTAATCTTAAATGATG 240
DB 195 TTATANTNGAANCGGCMNGAAGCAACCAACTTATGTCATTTATCCNNAANGAGG 254
QY 241 TGTTTTATGACTAATPACAGTATTTTTCATPAGAAACCATGTGTAATAATTTTAT 300
DB 255 NNTTTNANNACTAATNCCCNATTTTCCATANGAANCCNNNTTAAANTTTTNTAT 314
QY 301 TTTAAAAATAGCCGTGTTCAAGCTGATCATATTTCTTTATTTGATTTGGGAAGA 360
DB 315 TTTAAAAATPACCCNGTNTCCACCNGATCAAAATTCCTTNAATTTGATTTGGGAAA 374
QY 361 AAATACGTTTCTGATAGCATGAATGCAAAATTTTATGATTTTAAATCTGCTAATTTT 420
DB 375 AAATNCGTTCCNNAATCCNNGAANGGAANTTTTAAATTTTAAOCCCTANTTTT 434
QY 421 AAGAACTATTGAGAAATGATTAATGACATGACGACAACTAATTTACTGGCCAGCT 480
DB 435 AAAANCTATNGAANAANTGATTANNGACTTGAATTC-CAACCTANTTNCNGCCACN 493
QY 481 GTTGCAATGTTCTTACTTACTGTTCCCAAGAAAACCTTAACTGAATCTTACG 540
DB 494 GTGGCCTNGTNTCTTACTTANTCCCCCAAGAAANCTTAAACNGAANTCCNCC 553
QY 541 NGAAATACCTTAATATATCTTTAGCAAAAC--AAACTTTTGTGTTACATAGTTCT 598
DB 554 AAAATTAACCTTAATNTCTTGTGTAACAAANCAAAACGTTTNTGTTACANTNCT 613
QY 599 TTGATTTTACTGTTCTTATTTATTTCTGAACTCCATTTTCCCAAGACATTAATTAC 658
DB 614 TGGATTTTACGGGTCCCAATTTNATCCGAACCAATTTTCCCAACATANTTAC 673
QY 659 CCTATTAACCTTGTATNGACAGTGT 687
DB 674 CAT-TTTACTTGTGTAAGCAGTGT 701
```

```
RESULT 8
US-10-102-524-161/C
; Sequence 161, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-161
```

```
Query Match 23.6%; Score 161.8; DB 14; Length 510;
Best Local Similarity 94.8%; Pred. No. 5e-20; Indels 3; Gaps 3;
Matches 199; Conservative 0; Mismatches 8; Indels 3; Gaps 3;
```

```
QY 479 CTGTGGCAATGTGTTCTTACTTATGTTCTCCAGAGAAATCTTAACTGAATCTTCA 538
DB 510 CTGTGGCAATGTGTTCTTACTTATGTTCTCCAGAGAAATCTTAACTGAATCTTCA 451
QY 539 GCNGAATAA-CTTAATATATCTTGTATGCAAAACAAA-CTTTTGTGTTACATAGTT 596
```



Db 450 GCAGATAATCCTTAAATATCTTTGTAAAGCAAAACAAAGCTTTTTTGTTCATATGTT 391

Qy 597 CTTTGGATTTTACGTGTTCTCTAAATTTTATTTCTGAACCTCATTTTTCGCCAGACATAAT 656

Db 390 CTTTGGATTTTACTGTCTCTAATTTTATTTCTGAACTCAATTTTACCCGAGACCATAT 311

Qy 657 ACCCTATTTAACTTGTATGACAGCTGT 686

Db 330 ACCATA-TTAACTTTGTAAATGACAGCTGT 302

```

RESULT 9
US-10-102-524-188/C
; Sequence 188, Application US/10102524
; Publication No. US20030109434v1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITILE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102.524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-188

```

	Query Match	Similarity	Score	DB	Length
	Best Local	94.8%	Pred. No. 56-20;		510;
	Matches	199;	Conservative	0;	Mismatches 8; Indels 3; Gaps 3
Qy	479	CTGTGGCATTGTGTTCTTACTAGTCTCCAGGAAACCTTAACTGATCTCA	538		
Db	510	CTGTGGCATTGTGTTCTTACTAGTCTCCAGGAAACCTTAAATGATCTCA	451		
Qy	539	GCNGAATA-CCTTAAATATCTTTGTAGCCAAAGAAA-CTTTTTGTTCATAGT	556		
Db	450	GCAGATATTCCTTAAATATCTTTGTAGCAAAACAAAGCTTTTTGTTCATAGT	391		
Qy	557	CTTTGGATTTTACTGTTCCCTAATTTATTCGAACTCCATTTTTTCCAGACCATATT	656		
Db	390	CTTTGGATTTTACTGTTCCCTAATTTATTCGAACTCCATTTTACCCAGACCATATT	331		
Qy	657	ACCTTATTTAATCTTTGTATGCAAGTGT	686		
Db	330	ACCATAT-TTAACTTTGTATGCAAGTGT	302		

```

RESULT 10 US-10-102-524-645
; Sequence 645, Application US/10102524
; Publication No. US20030109434v1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT OF INFLAMMATORY DISEASES OF THE KIDNEY
; FILE REFERENCE: 210121.572
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FASTSEQ for Windows Version 4.0

```

```

; SEQ ID NO 645
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-645

```

Query Match	23.6%	Score 161.8	DB 14	Length 510
Best Local Similarity	94.8%	Pred. No. 56-20	8	Indels 3
Matches 199	0	Mismatches	3	Gaps 3

QY	479	CTGTGGGATGTGTTTCTTACTTAAGTGTCCCAAGGAAACCTTAAATCTTCA	538
Db	1	CTGTGGCATGTGTCTTCTTACTTAAGTGTCCCAAGGAAACCTTAAATCTTCA	60
QY	539	GCNGAATAA-CCTTAATAATACCTTGTAGCCAAACAAA-CTTTTTGTTACATAGTT	566
Db	61	GCAGAAATATCCTTAATAATACCTTGTAGCAAAACAAAAGCTTTTTGTTACATAGTT	120
QY	587	CTTTGGATTTTACTGTGCTAATTTTATTTCTGAACCTCATTTTTCCCGACCAATATT	656
Db	121	CTTTGGATTTTACTGTCTCTAATTTTATTTCTGAACCTCAATTTTACCCGACCAATATT	180
QY	657	ACCCATTTTAACTTGTGTAAGCAGAGTTGT	686
Db	181	ACCATTA-TTAACTTGTGTAAGCAGAGTTGT	209

RESULT 11  
US-09-960-352-5962  
; Sequence 5962, Application US/09960352

```

; PATENT NO: 052002013,15941
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5962
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 26-LIB188-014-01-E1-G5
; US-09-960-352-5962

```

Query Match	15.1%	Score 103.6;	DB 10;	Length 397;
Best Local Similarity	71.9%	Pred. No. 1.4e-09;		
Matches 192; Conservative	0;	Mismatches 65;	Indels 10;	Gaps 4;

QY 421 AAGACCTTTGGAAATGATTAAATGACATGAAAGCACAACACTAATTACGTGGCAGCT 480

Db 14 AATTATCCTGAGAAATATTATTAAGGTCACTTAAATGCAACAACATTATTAATCTGTCAGCT 73

QY 481 GTTGGCATGTGTCTTCTACTAGTCTCTCCCAAGMAA CTGTAACTGAATCTTCAGC 540

Db 74 AATAATGATAT-TCTGTCTCTTCATCTCTCCCAAGMAA CT-----TGAATTTTCGGC 125

QY 541 NGAATTAACCTTAATATATACCTTTGTAGCCNAACMAA CTTTGTGTTACATAGTCTTT 600

Db 126 AGAATTAATCTTCAAATATATACATATATTAAGCAAAATAGAGACTTCTGTTACATACCTT-TTT 184

QY 601 GGATTTTACTGTCCCTAAATTTTATTTCTGAAACTCTCATTTTTCGCCAGACATATTAACC 660

Db 185 GATTTTGTATTTCTTAACCTTATTTCTAATAACTCATATTTTACCCCAAAACATGATTACCA 244

QY 661 TATTTAACTTTGTATGCAAGTGTGT 687

Db 245 TA-TTAACCTTGTAAATGACAGTGTGT 270

RESULT 12  
US-10-240-485-151  
Sequence 151, Application US/10240485  
Publication No. US20030148327A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPERROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with  
TITLE OF INVENTION: Metastasis  
FILE REFERENCE: 5013.1007  
CURRENT APPLICATION NUMBER: US/10/240,485  
PRIOR FILING DATE: 2002-10-02  
PRIOR APPLICATION NUMBER: PCT/EP01/03970  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 202  
SEQ ID NO 151  
LENGTH: 7676  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-485-151

Query Match 9.7%; Score 66.8; DB 12; Length 7676;  
Best Local Similarity 48.5%; Pred. No. 0.016;  
Matches 210; Conservative 0; Mismatches 220; Indels 3; Gaps 1;  
QY 21 ATTCCTCCAAAGCATTATTTATCTATCTCAGTAATTTAANAATAAATCTT--- 77  
DB 3697 ATTTGTTAATATATATATATATATATATATGTTGTTAGTAATTTAATTTAATTTAATATATTA 3756  
QY 78 AGATTTAGAAAACTAGAGAAAAAGATNAATGAGATTAATTTAACTTACATGAAAAAGGA 137  
DB 3757 TATATTTATTTATTTATATATATATATATATAGATGAAATATTTAAGAAATTTGAAAGTTT 3816  
QY 138 AAATTTATACAAAGAGCTGAGAACCTTAAATTTGAATTTGATTAATTTGAAACTG 197  
DB 3817 AATGATGCTTAAATTTTATATATAGGAATATTTTATTAATAGATTAATTAAGAAATTT 3876  
QY 198 CATCGAAAGCAACTTATTTGTCATTTCTTAATGATGCGTTTATAGACTAATAC 257  
DB 3877 AATTTAATTTAAAAATTTTATGTTTATTTATGATTAAGATTTTATTTATTTAATTA 3936  
QY 258 ACTGATTTTCAATAGGAAACCCATGTTAAAAATTTTATTTTAAATTAAGCCGTG 317  
DB 3937 GATATTTTATTTTATTAAGAGCTTTAATTAATTTTATTTTATTTTATTTAATTA 3996  
QY 318 GTTCAAGCTCTGATCATTTCTTTATTTGATTTGGGAAAGAAATCTGTTCTGATA 377  
DB 3997 TTTTGT 4056  
QY 378 GCATGAATGCAAAATTTTATGATTTTAACTGCTAATTTTAAAGCTATGGAAT 437  
DB 4057 AAAATAGTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4116  
QY 438 TGATTAATGACAT 450  
DB 4117 AGTAGAATAGTAT 4129

RESULT 13  
US-10-311-455-2147/C  
Sequence 2147, Application US/10311455

Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPERROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311,455  
CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 2147  
LENGTH: 113515  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-2147

Query Match 9.5%; Score 65; DB 12; Length 113515;  
Best Local Similarity 44.8%; Pred. No. 0.083;  
Matches 282; Conservative 0; Mismatches 344; Indels 3; Gaps 1;  
QY 5 TAAATTTCTATTAACATTTCTTCAAGCATTTATTTATCTATCTCAGTAATTTT 64  
DB 45803 TAAATAATATCCAAATTTTATTTATTTAATCTCAAAAAATATATATATTTAATTTCA 45744  
QY 65 AANAATATACATTAATTAATTAAGAAAACTAGAAAAAGATNAATGAGATTAATTTAACTT 124  
DB 45743 AAAAAAATCCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 45684  
QY 125 ACATGAAAAAGAAATTTATTAACAAAGAGCTGAGAACCTTAATTTAATTTGAATTTGATTTAT 184  
DB 45683 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 45624  
QY 185 AATTTGAAACTGATCTGAAAGCAACTTATTTGTTCAATTTATTTATTTGATTTGGAAGAAAT 244  
DB 45623 AATCTAATCTAATAATTAATTTTCAATTAATTAATTAATTTATTTATTTCCAAATTAAC 45564  
QY 245 TTATGACTAATACCTGATTTTCAATTAAGAAACCATGTTAAAAATTTTATTTTAA 304  
DB 45563 ATCAAAATACATTC---CACTAAACAAATATATTTACATTTTCAATAACAAAAATTAATTT 45507  
QY 305 AAAATAGCCGTGCTGACGCTGATCATTTCTTTATTTGATTTGATTTGGAAGAAAT 364  
DB 45506 TCATTAATCAATTTTCTACATTTTCAAAATTAATTAATTAATTAATTAATTAATTA 45447  
QY 365 ACTGTTCTGATAGCATGAATGCAAAATTTTATTTTAACTGCTAATTTTAA 424  
DB 45446 ACAAACTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTTTATTTTCAAA 45387  
QY 425 ACTATGAGAAATGATTAATGACATGAGTGCACAACTAATTTACTGCGCCAGCTGTG 484  
DB 45386 TCTTTCAAAATTTTCTTAACTGATCAATCACTATTTTAACTAATTTTCAAAAAATTA 45327  
QY 485 GCATGCTGTTCTTACTAATGTTCTCCCAAGAAACCTTAATCTGATCTGCGGAA 544  
DB 45326 ACTAATAATTAATTAATTTTATTTTCTCAATTTATTTATTTATTTATTTATTTATTT 45267  
QY 545 TAACTTAATTAATTTGTTAGCAACAACTTTTGTTTTATTTTACTAGTTCTTTGAT 604  
DB 45266 TACATTTTAAACAATATTTCTATTAATCTATTAATTAATTAATTTATTTTATTTACT 45207  
QY 605 TTTACTGTTCTAATTTTATTTGAAACT 633  
DB 45206 TTAATAATTTTAAAAATATATCTTTATTT 45178

```
RESULT 14
US-10-198-846-6381/c
Sequence 6381, Application US/10198846
Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6381
LENGTH: 960
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1, 8, 66, 71, 79, 81, 84, 85, 86, 88, 117, 119, 125, 134,
LOCATION: 149, 156, 162, 166, 172, 179, 180, 194, 230, 249, 250, 253,
LOCATION: 268, 271, 273, 274, 275, 276, 277, 286, 291, 299, 312, 333,
LOCATION: 359, 366, 374, 393, 404, 406, 409, 411, 422, 424, 427
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 428, 433, 450, 465, 466, 468, 471, 487, 510, 512, 513, 516,
LOCATION: 517, 518, 528, 530, 534, 539, 553, 555, 559, 575, 580, 588,
LOCATION: 590, 610, 625, 626, 642, 645, 656, 658, 659, 675, 685, 686,
LOCATION: 690, 700, 702, 706, 716, 731, 745, 748, 756, 798, 802
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 807, 808, 819, 820, 830, 838, 840, 844, 845, 849, 861,
LOCATION: 871, 874, 875, 891, 892, 895, 898, 900, 901, 902, 904, 907,
LOCATION: 917, 923, 925, 926, 927, 929, 955
OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6381

Query Match 9.2%; Score 63.4; DB 14; Length 960;
Best Local Similarity 41.5%; Pred. No. 0.034;
Matches 199; Conservative 0; Mismatches 273; Indels 8; Gaps 1;

QY 53 TCACGATTTTAAATAAATTAACATTAGTATGAAAACTAGAAAAAGATNAATGAC 112
DB 930 TNNANNANACCTANACCTTTAAANATNNNNATNNANNAAAAAAATTAATAANNAN 871
QY 113 ATAATTAACCTACATGAAAAAGAAAAATTAACAAAGACCTAGAACTTAATAATG 172
DB 870 AAAATTTTNNAAAACTAATNTNANNTNANNTAATTTTAAAAAAANNNAAAAATA 811
QY 173 AATGACATTAATTAATTTGAAAACTGACATGAAAACTTAATTTGTCATTAATTCCT 232
DB 810 AANNNTTNTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 751
QY 233 AATGATGCTGTTTATGACATATACACTGATTTTCAATTAAGAAACCATGTAATAAT 292
DB 750 AANTNATATTTTATTAATNAAAAAAATTTNAATTTTAAAC-----AANTNT 699
QY 293 ATTTTATTTTAAAAATTAAGCTGCTCAAGCTCTGATCATATTTCTTTATTTGAT 352
DB 698 TTTTNTTNTTNNATNAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 639
QY 353 TGGAGAGAAATTAATGCTTCTGATGACATGAATGCAAAATTTTGAATTTTAACTCN 412
DB 638 TTTAAAAAATTAATTTTATTTTATTAATTAATTAATTAATTTATTTTCTTNA 579

RESULT 15
US-10-311-455-1445/c
Sequence 1445, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
nition of Cytosine Methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1445
LENGTH: 6106
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1, 8, 66, 71, 79, 81, 84, 85, 86, 88, 117, 119, 125, 134,
LOCATION: 149, 156, 162, 166, 172, 179, 180, 194, 230, 249, 250, 253,
LOCATION: 268, 271, 273, 274, 275, 276, 277, 286, 291, 299, 312, 333,
LOCATION: 359, 366, 374, 393, 404, 406, 409, 411, 422, 424, 427
OTHER INFORMATION: n = A,T,C or G
US-10-311-455-1445

Query Match 9.2%; Score 63; DB 12; Length 6106;
Best Local Similarity 46.5%; Pred. No. 0.073;
Matches 198; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

QY 38 ATTTATCTATATCTCAGTAAATTTTAAANAATTAACATTAGTATGAAAACTAGAA 97
DB 3138 ATTCAATACATTAATCTCAATTAACCAATATATTAATTAATTAATTAATTAATTAAT 3079
QY 98 AAAAGATNAATGACATTAATTAACCTTACATGAAAAAGAAAAATTAACAAAGACTGA 157
DB 3078 AAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3019
QY 218 TGTTCATTAATTTCTATGATGAGTGTGTTTATGACTATATGACGATTTTCAATAGAA 277
DB 2958 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2899
QY 278 ACCCATGTAATAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 337
DB 2898 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2839
QY 338 TCTTTATTTTGAATTTGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 397
DB 2838 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2779
QY 398 AGATTTTAATTCNCATTAATTTTAAAGAACTATTAATTAATTAATTAATTAATTAATTAAT 457
DB 2778 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2719
QY 458 ACAACA 463
DB 2718 ATATA 2713
```

Tue Dec 2 06:11:50 2003

us-09-835-992a-20.rnpb

Page 10

Search completed: November 27, 2003, 12:52:57  
Job time : 263.057 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 06:00:22 ; Search time 50.7082 Seconds

(without alignments)  
5979.910 Million cell updates/sec

Title: US-09-835-992A-20

Perfect score: 687

Sequence: 1 attcaattcttcatcctaac.....cttctgtaagcagctgttc 687

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*\n2: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*\n3: /cgn2\_6/ptodata/1/lna/6A\_COMB.seq:\*\n4: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:\*\n5: /cgn2\_6/ptodata/1/lna/6C\_COMB.seq:\*\n6: /cgn2\_6/ptodata/1/lna/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	683	99.4	687	3	US-08-896-164-20
2	611.8	89.1	714	3	US-08-896-164-19
3	411.8	59.9	843	3	US-08-896-164-22
4	161	23.4	173	4	US-09-313-294A-1553
5	66.4	9.7	5852	1	US-07-867-106-2
6	56	8.2	20674	4	US-09-641-638-651
7	53.6	7.8	837	3	US-08-998-416-288
8	53.6	7.8	927	4	US-09-134-001C-150
9	51.8	7.5	615	3	US-08-998-416-186
10	51.4	7.5	6243	2	US-09-056-075-1
11	51.2	7.5	615	3	US-08-998-416-186
12	51	7.4	1692	4	US-09-601-198-63
13	50.2	7.3	6124	4	US-08-213-419B-3
14	49.2	7.2	636	3	US-08-998-416-1137
15	48.6	7.1	19124	2	US-08-487-826B-13
16	48.4	7.0	1511	1	US-07-991-867B-8
17	48.4	7.0	1511	1	US-08-107-755A-8
18	48.4	7.0	1511	2	US-08-544-332-8
19	48.4	7.0	1511	4	US-09-370-861A-8
20	48	7.0	1431	3	US-09-316-083-2
21	48	7.0	1431	4	US-09-933-700-2
22	48	7.0	20674	4	US-09-641-638-651
23	47.8	7.0	10640	4	US-09-417-485D-5
24	47.2	6.9	19124	2	US-08-487-826B-13
25	46.4	6.8	3232	4	US-09-333-214-3
26	46	6.7	5340	4	US-09-627-122-21
27	45.8	6.7	2251	3	US-08-991-677-11

C	28	45.8	6.7	6152	3	US-08-973-462-1	Sequence 1, Appl
C	29	45.4	6.6	5181	1	US-08-257-073-10	Sequence 10, Appl
C	30	45.2	6.6	658	3	US-08-998-416-595	Sequence 55, App
C	31	45.2	6.6	678	1	US-07-991-867B-23	Sequence 23, Appl
C	32	45.2	6.6	678	1	US-08-107-755A-23	Sequence 23, Appl
C	33	45.2	6.6	678	2	US-08-544-332-23	Sequence 23, Appl
C	34	45.2	6.6	678	4	US-09-370-861A-23	Sequence 23, Appl
C	35	45.2	6.6	945	4	US-09-601-198-177	Sequence 17, App
C	36	45.2	6.6	4185	4	US-09-417-485D-7	Sequence 7, Appl
C	37	45.2	6.6	6768	1	US-08-107-755A-1	Sequence 1, Appl
C	38	45.2	6.6	8457	1	US-07-991-867B-1	Sequence 1, Appl
C	39	45.2	6.6	8457	2	US-08-544-332-1	Sequence 1, Appl
C	40	45.2	6.6	8457	4	US-09-370-861A-1	Sequence 1, Appl
C	41	44.8	6.5	2539	3	US-08-749-522-3	Sequence 3, Appl
C	42	44.8	6.5	10640	4	US-09-417-485D-5	Sequence 5, Appl
C	43	44.4	6.5	1500	4	US-09-601-198-36	Sequence 36, Appl
C	44	44.4	6.5	15016	4	US-09-601-198-60	Sequence 60, Appl
C	45	44.2	6.4	1939	1	US-07-715-751B-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-896-164-20  
Sequence 20, Application US/08896164  
Patent No. 6218521  
GENERAL INFORMATION:  
APPLICANT: OBATA, Yuichi  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND METHODS FOR  
TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/896,164  
FILING DATE: July 17, 1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6218521man D. Hanson  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 687 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-896-164-20  
Query Match 99.4%; Score 683; DB 3; Length 687;  
Best Local Similarity 100.0%; Pred. No. 1.5e-129;  
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 ATTTAATTTTCCTATTAAACATTCTTCTCAAGACATATTTTATCTCATATCTGACGAA 60  
Db 1 ATTTAATTTTCCTATTAAACATTCTTCTCAAGACATATTTTATCTCATATCTGACGAA 60  
Cy 61 TTTTAAATAATTAACATAGTATTAGAAAACTGGAAGAAAGATMAATGCAATTAATTA 120

```

Db      61 TTTTAAATAATACATTGATGATTAAGAAAACTAGAAAAAAGATTAATGACATATTTAA 120
Qy      121 ACTTACATGAAAAAGAAAAATTTATACAAAGACCTGAGACGTTATTAATGAAATGGA 180
Db      121 ACTTACATGAAAAAGAAAAATTTATACAAAGACCTGAGACGTTATTAATGAAATGGA 180
Qy      181 TTATATTTGAAAACTGATCGTGAAGAAACCTTATGTTCAATATTTCTTAATGANG 240
Db      181 TTATATTTGAAAACTGATCGTGAAGAAACCTTATGTTCAATATTTCTTAATGANG 240
Qy      241 TGTTTATGACTAATACATGATTTTTCATATAGGAAACCATGTTAAATAATTTTAT 300
Db      241 TGTTTATGACTAATACATGATTTTTCATATAGGAAACCATGTTAAATAATTTTAT 300
Qy      301 TTTTAAATAATGACCTGCTGTTCAAGCTCGATCATATTTCTTTTATTTGATTTGGGA 360
Db      301 TTTTAAATAATGACCTGCTGTTCAAGCTCGATCATATTTCTTTTATTTGATTTGGGA 360
Qy      361 AAATACGTTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420
Db      361 AAATACGTTTCTGATGACATGAAATGCAAAATTTTATGATTTTATCTGCTAATTTT 420
Qy      421 AAGACATTTGAGAAATGATTAATGACATGAGTGACACAACATTAATTAATGAGCAGCT 480
Db      421 AAGACATTTGAGAAATGATTAATGACATGAGTGACACAACATTAATTAATGAGCAGCT 480
Qy      481 GTTGCAATGTTGTTCTTACTTACTTACTGTTCTCCAGAGAAACCTTTAACTGAATCTTCAGC 540
Db      481 GTTGCAATGTTGTTCTTACTTACTTACTGTTCTCCAGAGAAACCTTTAACTGAATCTTCAGC 540
Qy      541 NGATTAACCTTAATATATCTTTGTTAGCCAAACAACTTTTGTGTTACATGATGCTTT 600
Db      541 NGATTAACCTTAATATATCTTTGTTAGCCAAACAACTTTTGTGTTACATGATGCTTT 600
Qy      601 GGATTTTACTGTTCTTCTAATTTTATTCGAAACCTCAATTTTCCCGAGACATTAATTAACC 660
Db      601 GGATTTTACTGTTCTTCTAATTTTATTCGAAACCTCAATTTTCCCGAGACATTAATTAACC 660
Qy      661 TATTTTACTTTGTTATGACAGTTGTT 687
Db      661 TATTTTACTTTGTTATGACAGTTGTT 687

```

RESULT 2  
US-08-896-164-19  
; Sequence 19, Application US/08896164  
; Patent No. 6218521  
; GENERAL INFORMATION:  
; APPLICANT: OBATA, Yuichi  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED  
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR  
; NUMBER OF SEQUENCES: 87  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/896,164  
; FILING DATE: July 17, 1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6218521man D. Hanson  
; REGISTRATION NUMBER: 30,946

```

; REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-896-164-19

```

Query Match 89.1%; Score 611.8; DB 3; Length 714;  
Best Local Similarity 97.1%; Pred. No. 3.4e-115;  
Matches 671; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

```

Qy      1 ATTTTAAATTTTCTATTTAAACATTTCTCAAGCATTATTTATCTATATCTCACTGAA 60
Db      15 ATTTTAAATTTTCTATTTAAACATTTCTCTCAAGCATTATTTATCTATATCTCACTGAA 74
Qy      61 TTTTAAATAATACATTAGTTTNGAAAAAACTAGAAAAAAGATTAATGACATATTTAA 120
Db      75 TTTTAAATAATACATTAGTTTNGAAAAAACTAGAAAAAAGATTAATGACATATTTAA 134
Qy      121 ACTTACATGAAAAAGAAAAATTTATTAACAAGACCTGAGAACTGATTAATTTGAAATGAGA 180
Db      135 ACTTACATGAAAAAGAAAAATTTATTAACAAGACCTGAGAACTGATTAATTTGAAATGAGA 194
Qy      181 TTATATTTTGAATACTGATCTGAAAGCAAACTTTATGTTGCAATTAATCTTAATGATGG 240
Db      195 TTATATTTTGAATACTGATCTGAAAGCAAACTTTATGTTGCAATTAATCTTAATGATGG 254
Qy      241 TGTTTATGACTAATACATGATTTTTCATTAAGAAAAACCATGTTAAATAATTTTAT 300
Db      255 TGTTTATGACTAATACATGATTTTTCATTAAGAAAAACCATGTTAAATAATTTTAT 314
Qy      301 TTTTAAATAATGACCTGCTGTTCAAGCTCGATCATATTTCTTTTATTTGATTTGGGAAGA 360
Db      315 TTTTAAATAATGACCTGCTGTTCAAGCTCGATCATATTTCTTTTATTTGATTTGGGAANA 374
Qy      361 AAATACGTTTCTGATGACATGAAATGCAAAATTTTATGATTTTAAATCTGCTAATTTT 420
Db      375 AAATACGTTTCTGATGACATGAAATGCAAAATTTTATGATTTTAAATCTGCTAATTTT 434
Qy      421 AAGACATTTGAGAAATGATTAATGACATGAGTGACACAACATTAATTAATGAGCAGCT 480
Db      435 AAGACATTTGAGAAATGATTAATGACATGAGTGACACAACATTAATTAATGAGCAGCT 494
Qy      481 GTTGCAATGTTGTTCTTACTTACTTACTGTTCTCCAGAGAAACCTTTAACTGAATCTTCAG 539
Db      495 GTTGCAATGTTGTTCTTACTTACTTACTGTTCTCCAGAGAAACCTTTAACTGAATCTTCAG 554
Qy      540 CNGAATAA-CCTTAAATATACCTTTGTTAGCCAAACAAA-CCTTTTGTGTTACATAGTTC 597
Db      555 CAGATATATCCTTAAATATACCTTTGTTAGCCAAACAAAACCTTTTGTGTTACATAGTTC 614
Qy      598 TTT-GCATTTTACTGTTCTTAAATTTTATTTCTGAAACCTCAATTTTCCCGAGACATTAAT 656
Db      615 TTTGGGATTTTACTGTTCTTAAATTTTATTTCTGAAACCTCAATTTTCCCGAGACATTAAT 674
Qy      657 ACCCTATTTAATCTGTTATGACAGTTGTT 687
Db      675 ACCATA-TTAACTTTGTTATGACAGTTGTT 704

```

RESULT 3  
US-08-896-164-22  
; Sequence 22, Application US/08896164  
; Patent No. 6218521  
; GENERAL INFORMATION:  
; APPLICANT: OBATA, Yuichi  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED  
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR

```

; TITLE OF INVENTION:  DIAGNOSING AND TREATING GASTRIC CANCER
; NUMBER OF SEQUENCES:  87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Felte & Lynch
; STREET:  805 Third Avenue
; CITY:  New York City
; STATE:  New York
; COUNTRY:  USA
; ZIP:  10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Diskette, 3.50 inch, 1.44mb
; COMPUTER:  IBM PS/2
; OPERATING SYSTEM:  PC-DOS
; SOFTWARE:  Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/896,164
; FILING DATE:  July 17, 1997
; CLASSIFICATION:  424
; ATTORNEY/AGENT INFORMATION:
; NAME:  NO. 6218521man D. Hanson
; REGISTRATION NUMBER:  30,946
; REFERENCE/DOCKET NUMBER:  LUD 5499 - JEL/NDH/SLH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (212) 688-9200
; TELEFAX:  (212) 838-3884
; INFORMATION FOR SEQ ID NO:  22:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  843 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  double
; TOPOLOGY:  linear
;
; US-08-896-164-22
;
Query Match      59.9%; Score 411.8; DB 3; Length 843;
Best Local Similarity 71.8%; Pred. No. 6.8e-75;
Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;
;
Qy 1 ATTTTATTTTCTATTAACATTTTCTCAAGACATTTATTTATCTATATCTCACTGAA 60
Db 15 ATTTTATTTTCTATTAACATTTTCTCAAGACATTTATTTATCTATATCTCACTGAA 74
Qy 61 TTTTAAATAAATCAATAGTATTAAGAAAAGTAAAGTAAAGTAAAGTAAAGTAA 120
Db 75 TTTTAAATAAATCAATAGTATTAAGAAAAGTAAAGTAAAGTAAAGTAAAGTAA 134
Qy 121 ACTTACATGAAAAGAAAATTTATACAAAGACTGAAAGCTTATTAATGAAATGAA 180
Db 135 CTTTNCCTGAAAAGAAAATTTATACAAAGACGAAAGAAAGTAAATGAAATGAA 194
Qy 181 TTATTAATTTGAAAAGCTGATCTGAAAGCAAACTTTATTTGTTCAATTAATTTAA 240
Db 195 TTATTAATTTGAAAAGCTGATCTGAAAGCAAACTTTATTTGTTCAATTAATTTAA 254
Qy 241 TGTATTATGACTATACATGATTTTTCATTAAGAAACCATGTTAAATAATTTTAT 300
Db 255 NNTTTANNNACTAATNCCNGATTTTCCATTAAGAAACCCNNNTTAAATNTTTTAT 314
Qy 301 TTTTAAATAAAGCTGTGTTCAAGCTGTGATCATATTTCTTTTATTTTGAATTTGAGA 360
Db 315 TTTTAAATAAAGCTGTGTTCAAGCTGTGATCATATTTCTTTTATTTTGAATTTGAGA 374
Qy 361 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTGAATTTTAAATCTGCTAATTTT 420
Db 375 AAATACGTGTTCTGATAGCATGAAATGCAAAATTTTGAATTTTAAATCTGCTAATTTT 434
Qy 421 AAGAATCTAGGAAATGATTAATGACATGAAGTGAACCACTAATTTACTGCGCACT 480
Db 435 AAAAATCTATGAAAATGATTAATGACATGAATTTGCGCACTAATTTACTGCGCACT 493
Qy 481 GTTGCAATGTGTTTCTTACTAGTTCTCCCAAGAAAACCTTAAACTGAATCTTGAGC 540
Db 494 GTTGCAATGTGTTTCTTACTAGTTCTCCCAAGAAAACCTTAAACTGAATCTTGAGC 553

```

```

Qy 541 NGATAACCTTAATATATCTTTGTAAGCAAC--AAAATTTTGTGTTATACATGTTCT 598
Db 554 AAAATACCTTAATATATCTTTGTAAGCAAC--AAAATTTTGTGTTATATATCTTCT 613
Qy 599 TTGATTTTACGTGTTCTTAATTTTATCTGAAATCTCAATTTTCCCGACATTAATTA 658
Db 614 TGGATTTTACGGGTGTTCTTAATTTTATCTGAAATCTCAATTTTCCCGACATTAATTA 673
Qy 659 CCTATTTTACTTTGTTATGACAGTTGTT 687
Db 674 CAT-TTTACCTGTTAAGGCNCAATGTT 701
;
RESULT 4
US-09-313-294A-1553/C
; Sequence 1553, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT:  Talgudi, Raghunath V.
; APPLICANT:  Ito, Laura Y.
; APPLICANT:  Sherman, Bradley K.
; TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE:  PL-0017 US
; CURRENT APPLICATION NUMBER:  US/09/313,294A
; CURRENT FILING DATE:  1999-05-14
; NUMBER OF SEQ ID NOS:  7600
; SOFTWARE:  PERL Program
; SEQ ID NO 1553
; LENGTH:  173
; TYPE:  DNA
; ORGANISM:  Zea mays
; FEATURE:
; NAME/KEY:  misc_feature
; OTHER INFORMATION:  Incyte ID No. 6476212 700551132H1
;
; US-09-313-294A-1553
;
Query Match      23.4%; Score 161; DB 4; Length 173;
Best Local Similarity 98.8%; Pred. No. 2e-24;
Matches 161; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
Qy 3 TTTTATTTTCTATTAACATTTCTTCGAAGCATTTATTTATCTATATCTCACTGAAT 62
Db 163 TTTTATTTTCTATTAACATTTCTTCGAAGCATTTATTTATCTATATCTCACTGAAT 104
Qy 63 TTAATAATAATCACTAGTATTAAGAAAAGTAAAGTAAAGTAAAGTAAAGTAAATTAAC 122
Db 103 TTAAGAATAATCACTAGTATTAAGAAAAGTAAAGTAAAGTAAAGTAAATTAAC 44
Qy 123 TTACATGAAAAGAAAATTTATCAAAAGACTGAGAACCTTA 165
Db 43 TTACATGAAAAGAAAATTTATCAAAAGACTGAGAACCTTA 1
;
RESULT 5
US-07-867-106-2/C
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT:  Slade, Martin B
; APPLICANT:  Chang, Andy C M
; APPLICANT:  Williams, Kelch L
; TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES:  19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls
; STREET:  One Liberty Place 46th Floor
; CITY:  Philadelphia
; STATE:  PA
; COUNTRY:  USA
; ZIP:  19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk

```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
NAME/KEY: CDS
LOCATION: 2378..5038
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 9.7%; Score 66.4; DB 1; Length 5852;
Best Local Similarity 46.8%; Pred. No. 3.2e-05;
Matches 240; Conservative 0; Mismatches 269; Indels 4; Gaps 1;

QY 2 TTTTAAATTTCTATTAACATCTTCCAAAGCATTTATTTATCTCATCTGAT 61
DB 5751 TTGTTATTTTATTAATGTTATTTGTTGTTTCTTATTAATTTCTATTTTAA 5662
QY 62 TTTAANAATAACATTAGTATTAGAAAAAAGTGAAGAAAAAGTAAATGACATATTTAA 121
DB 5691 TTTTAAATTAATTAATTAATTTTAAATAAATAAAAAAATTTTAA 5662
QY 122 CTTACATGAAAAAGAAATTTATACAAAGACGAGACCTTAAATTTGAATGAT 181
DB 5631 ATTAAAAAATTTAAAAAATTTAAAAAATTTAAAAATTTGAAATTTAAA 5572
QY 182 TATTAATTTGAAAACTGATCTGAAAGCAACTTATTTGTTCAATTTCTTAATGATG 241
DB 5571 TAAAAAATGTAAGGGTTTCTTTTAAATTAATGATGATTTTATTTTAAATCATTTGA 5512
QY 242 GTTTTATGACTAATACATGATTTTTCATTAAGGAAACCATGTTAAAAATTTTAT 301
DB 5511 CGAGATTTAAAAAATCTTAAAAACATTAATTTGATTTTCTTTTCTTTTCTTTT 5452
QY 302 TT---AAAAATTAAGCTGTTGTTCAAGCTGATCATTTCTTTTATTTGATTGGGA 357
DB 5451 TTTTAAAAAATTTCAATTAATTTAAATTTATCTTAATCTTGATGAACTTCAATTTT 5392
QY 358 AGAAAAATACGTTTCTGATGACATGAAAGTGAATTTTAAATTTTAAATCTGAT 417
DB 5391 TAAATTTTAAATTAATTTTAAATTTTAAATTTGATGATCTAATTAATTTTCCATG 5332
QY 418 TTTAAGACTATTGAGAAATTTGATTAATGACATGAAAGTGAACATTAATTTACTGGCA 477
DB 5331 TTTTAAATTTTCTTTTAAAGTATTTTAAATTAATTAATTAATTAATTAATTAATGA 5272
QY 478 GCTGTTGGCAATGTTGTTCTTACTTAAGTTCTCC 510
DB 5271 AATGATACCTAGATCTTTTCCCAAGTTTCC 5239
```

```
RESULT 6
US-09-641-638-651/C
Sequence 651, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bouguetelert, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Amick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GENSET.05ICP1
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
```



LOCATION:	17063..17554	
OTHER INFORMATION:	exon 14	
NAME/KEY:	misc_feature	
LOCATION:	17555..20674	
OTHER INFORMATION:	3 regulatory region	
NAME/KEY:	allele	
LOCATION:	1128	
OTHER INFORMATION:	10-508-191	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	1182	
OTHER INFORMATION:	10-508-245	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	1559	
OTHER INFORMATION:	10-509-284	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	1570	
OTHER INFORMATION:	10-509-295	: deletion of C
NAME/KEY:	allele	
LOCATION:	1827	
OTHER INFORMATION:	10-510-173	: variable motif ATTGA or TTTT
NAME/KEY:	allele	
LOCATION:	2048	
OTHER INFORMATION:	10-511-62	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	2323	
OTHER INFORMATION:	10-511-337	: insertion of T
NAME/KEY:	allele	
LOCATION:	2341	
OTHER INFORMATION:	10-512-36	: polymorphic base G or C
NAME/KEY:	allele	
LOCATION:	2623	
OTHER INFORMATION:	10-512-318	: polymorphic base A or G
NAME/KEY:	allele	
LOCATION:	2832	
OTHER INFORMATION:	10-513-250	: polymorphic base A or G
NAME/KEY:	allele	
LOCATION:	2844	
OTHER INFORMATION:	10-513-262	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	2934	
OTHER INFORMATION:	10-513-352	: polymorphic base A or G
NAME/KEY:	allele	
LOCATION:	2947	
OTHER INFORMATION:	10-513-365	: polymorphic base A or G
NAME/KEY:	allele	
LOCATION:	3802	
OTHER INFORMATION:	12-206-81	: polymorphic base A or G
NAME/KEY:	allele	
LOCATION:	4062	
OTHER INFORMATION:	10-343-231	: deletion of C
NAME/KEY:	allele	
LOCATION:	4088	
OTHER INFORMATION:	12-206-366	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	4109	
OTHER INFORMATION:	10-343-278	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	4170	
OTHER INFORMATION:	10-343-339	: polymorphic base G or T
NAME/KEY:	allele	
LOCATION:	5903	
OTHER INFORMATION:	10-346-23	: polymorphic base A or G
NAME/KEY:	allele	
LOCATION:	6019	
OTHER INFORMATION:	10-346-141	: polymorphic base A or G
NAME/KEY:	allele	
LOCATION:	6141	
OTHER INFORMATION:	10-346-263	: polymorphic base G or C
NAME/KEY:	allele	
LOCATION:	6183	
OTHER INFORMATION:	10-346-305	: polymorphic base C or T
NAME/KEY:	allele	
LOCATION:	6338	

	OTHER INFORMATION:	10-347-74	:	polymorphic base A or G
	NAME/KEY:	allele		
Dd	LOCATION:	6375		
	OTHER INFORMATION:	10-347-111	:	polymorphic base G or C
	NAME/KEY:	allele		
	LOCATION:	6429		
	OTHER INFORMATION:	10-347-165	:	polymorphic base C or T
	NAME/KEY:	allele		
	LOCATION:	6467		
	OTHER INFORMATION:	10-347-203	:	polymorphic base A or G
	NAME/KEY:	allele		
	LOCATION:	6484		
	OTHER INFORMATION:	10-347-220	:	polymorphic base A or G
	NAME/KEY:	allele		
	LOCATION:	6534		
	OTHER INFORMATION:	10-347-271	:	polymorphic base A or T
	NAME/KEY:	allele		
	LOCATION:	6611		
	OTHER INFORMATION:	10-347-348	:	polymorphic base A or G
	NAME/KEY:	allele		
	LOCATION:	7668		
	OTHER INFORMATION:	10-348-391	:	polymorphic base A or G
	NAME/KEY:	allele		
	LOCATION:	8608		
	OTHER INFORMATION:	10-349-47	:	polymorphic base C or T
	NAME/KEY:	allele		
	LOCATION:	8658		
	OTHER INFORMATION:	10-349-97	:	polymorphic base A or G
	NAME/KEY:	allele		
	LOCATION:	8703		
	OTHER INFORMATION:	10-349-142	:	polymorphic base G or C
	NAME/KEY:	allele		
	LOCATION:	8777		
	OTHER INFORMATION:	10-349-216	:	detection of CTG
	NAME/KEY:	allele		
	LOCATION:	8785		
	OTHER INFORMATION:	10-349-224	:	polymorphic base G or T
	NAME/KEY:	allele		
	LOCATION:	8926		
	OTHER INFORMATION:	10-349-368	:	polymorphic base C or T
	NAME/KEY:	allele		
	LOCATION:	12171		
	OTHER INFORMATION:	10-350-72	:	polymorphic base C or T
	NAME/KEY:	allele		
	LOCATION:	12429		
	OTHER INFORMATION:	10-350-332	:	polymorphic base C or T
	NAME/KEY:	allele		
	LOCATION:	13341		
	OTHER INFORMATION:	10-507-170	:	polymorphic base A or G
	NAME/KEY:	allele		
	LOCATION:	13492		
	OTHER INFORMATION:	10-507-321	:	polymorphic base A or C
	NAME/KEY:	allele		
	LOCATION:	13524		
	OTHER INFORMATION:	10-507-353	:	polymorphic base C or T
	NAME/KEY:	allele		
	LOCATION:	13535		
	Query Match	8.2%;	Score 56;	DB 4; Length 20674;
	Best Local Similarity	47.6%;	Pred. No.	0.0046;
	Matches 233;	Conservative	0;	Mismatches 243; Indels 13; Gaps 2;
Oy	TTCCTATTAAACATTTCTTGCAAGCATTAATTATCCATAATCTCAGTGAATTTAANA	68		
Dd	11528 TTAAAAATTTAAATTTAAAATTTAAATTAATTAATTAATTAATTTAAATTTAAAA	114659		
Oy	AATPAACAATTAGATTAGA AAAAACCTAGAAAAAAGATNAATGCAGTAATTAACCTACT	128		
Dd	11468 TTTTAATTTAATTTAATTTAAAATTTAAAAATTTAAATTTAATTTAATTTAATTTAAAA	11409		
Oy	129 GA AAAGAAAATTTATPACAAAGAGCTGGAAACGTATTAATTGAATGAGATTAAATT	188		
Dd	11408 TTAAAAATTTAATTTAATTTAATTTAA-----ATTAAATTTAAATTTAATTTAATTT	11360		

OY	189	TGAAACGCATCGAAGCAACTTATTGTTCATTTATCTTAATGAAGGGTTTTAT	248
Db	11359	AAAATTAAAAATATTTAAATTTTAAATTTTAAATTTAAATTTAAATTTAAAT	113000
OY	249	GACTAATPACATCATTTTTCATTAAGAAACCAGTGTTAAAAATATTTTAT--TTTAA	306
Db	11299	ATTAAATTTAAATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATA	11240
OY	307	AATAAGCCTGGTTCAGCCCTGATCATTTTCTTTATTTTGAATTTGGAGAAATTC	366
Db	11239	ATTTAAATGTTTATTATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAA	111800
OY	367	TGTTTCATAGCATGCAATATGCAAAATTTTATAGATTTTAAATCTCNCATTTTAAAGAC	426
Db	11179	TTATTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT	11120
OY	427	TATTGAGAAATGTATTATATGACATGAAGTGACACAACATAATTACTGGCACGTGTGC	486
Db	11119	AAATTTAAATTTTATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA	11060
OY	487	ATTGTGTTT 495	
Db	11059	ATGCTCCTT 11051	

  

```

RESULT 7
US-08-998-416-288/C
Sequence 288, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPILII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGCI976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 288:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
US-08-998-416-288

Query Match      7.8%; Score 53.6; DB 3; Length 837;
Best Local Similarity 45.6%; Pred. No. 0.01;
Matches 260; Conservative 0; Mismatches 303; Indels 7; Gaps 2;

QY      1  ATTTTAATTTTCTATTAATAACATCTTCTCAAGCAATATTTATCTATATCTCACTGAA 60
DB      569  AATTAAATAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 510
QY      61  TTTTAATAATAATACATTTAGTATTAAGAAAACCTAGGAAAAAGATNAATGACATATTA 120
DB      509  TAAATTAATAATAATTAATTAATTAAGAAATTAAGTTAAATTAATTTTATTAATTAATTTCTAT 450
QY      121  ACTTACATGAAAAAGAAAAATTATTAACAAGAGCTAGAACCTTAATTAATTTGAAATGAGA 180
DB      449  AAAAAGATTAATAATATATTAATCAACATATATTTATTAATAAATAGATATTAATAATAAAA 390
QY      181  TTAATATTTGAAAACTGCATCTGAAAGCAAACTTTATTTGTCATATTTCTTAATGATGG 240
DB      389  TAAATATTTACAAATTTTAAATTAATTAATTTTATTAATAATAATAATATATTTATTTAA 330
QY      241  TGTTTATGACCTAATACACATGATTTTCAATAGAGAAACCCATGTAATAAATATTTTAT 300
DB      329  TAAACAATTAATTAATA---TAATATATTAATTAATTAATTAATTAATTAATTAATTAATA 274
QY      301  TTTAAAAATAAGCCCTGTGTCAGCTCGATCAATATTTCTTTATTTGATTTGGAGAA 360
DB      273  AGAAAAATAATATATATCTATATAATATTTTAATACATTAATTTAAATTTGAACATAGACTAA 214
QY      361  AAATACGTGTTCTGATAGACATGAATGCAAAATTTTGAATTTTAACTCNCATATTTT 420
DB      213  ATAGATATCATATTAATAATATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATA 154
QY      421  AAGAACTTTGAGAAATTTGATTAATGACATGAAGGCACA---ACACTAATTAATCGGCCA 477
DB      153  GTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 94
QY      478  GCTGTGGCATTTGTTTCTTACTAGTGTCTCCCAAGAAAACTCTTAACATGAATCTTC 537
DB      93  ATGATATAATAATAGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 34
QY      538  AGCNGAATTAACCTTAATATATTAATCTTGTAG 567
DB      33  ATCTTATTAATTAATTAATTAATTAATTAATTTAG 4

RESULT 8
US-09-134-001C-150/c
; Sequence 150, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 150
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-150

Query Match      7.8%; Score 53.6; DB 4; Length 927;

```

Best Local Similarity 53.3%; Pred. No. 0.01;  
Matches 136; Conservative 0; Mismatches 115; Indels 4; Gaps 1;

QY 200 TCTGAAGCAAACTTATGTCATTAATCTTAAGATGCTGTTTATGATCAATACAC 259  
DB 606 TATTAATAGACTTATTTTCGTTTATATCTGTTTAAAGTCTCTTTGAATAAATCAC 547  
QY 260 TGAATTTTCAATAGAAACCCATGTTAAATTTTATTTTAAATAAGCTGTGT 319  
DB 546 TAAATTTTCAATAGAAATTTTGGCAAAATGATCTTTTACGTTAAATTTCCAGATAT 487  
QY 320 TCAAGCTGATCATATTTCTTTATTTTGGAGAAATATCT---GTTTGA 375  
DB 486 AAAAGCAATCATTTCTCTATTTTACATGTCATTAATTAATTAATATCTGT 427  
QY 376 TACATGAAATGCAAAATTTTATGATTTTAACTCCTAATTTTAAAGATTTGAA 435  
DB 426 TTTTAAATTAATTTGAAATTTTGGATTTTCTCTTATTAATCAATTAATTTAGGTA 367  
QY 436 ATTGATTAATGACT 450  
DB 366 ATGAACTGACACTAT 352

## RESULT 9

US-08-998-416-186  
; Sequence 186, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reibschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CCG1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8689  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 186:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:

ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 7.5%; Score 51.8; DB 3; Length 615;  
Best Local Similarity 45.9%; Pred. No. 0.023;  
Matches 213; Conservative 0; Mismatches 245; Indels 6; Gaps 1;

QY 6 AATTTCTAATTAACATTCCTCAAGCATTTTATGCTTATCTATCTACGATTTTA 65  
DB 42 ACTTTTATTAATTAATTTTAAGTATTAATTTTAACTATTTATTCATTTAAT 101  
QY 66 ANAATTAACATTAATTAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 125  
DB 102 AATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 161  
QY 126 CATGAAGAAAGAAATTAATTAAGAAAGTAAAGTAAAGTAAAGTAAAGTAA 185  
DB 162 TCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 221  
QY 186 ATTGAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 245  
DB 222 TGTTCAA-----ATTGAATTAATTAATTAATTAATTAATTAATTAATTA 275  
QY 246 TATGACTAATACATGATTTTCAATTAAGAAAGTAAAGTAAAGTAAAGTAA 305  
DB 276 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 335  
QY 306 AATTAAGCTGCTGCTCAAGCTGATCAATATTTCTTTATTTGGAAGAAATA 365  
DB 336 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 395  
QY 366 CTGTTCTGATGATCAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 425  
DB 396 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 455  
QY 426 CTATTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 469  
DB 456 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 499

## RESULT 10

US-09-056-075-1  
; Sequence 1, Application US/09056075  
; Patent No. 5955368  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Eric A.  
; APPLICANT: Bradshaw, Marilee  
; APPLICANT: Rood, Julian  
; TITLE OF INVENTION: Expression System for Clostridium  
; TITLE OF INVENTION: Species  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,075  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.95238  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3770..4013  
OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from  
OTHER INFORMATION: plasmid RP4"  
US-09-056-075-1

Query Match 7.5%; Score 51.4; DB 2; Length 6243;  
Best Local Similarity 46.8%; Pred. No. 0.034;  
Matches 191; Conservative 0; Mismatches 214; Indels 3; Gaps 1;

QY 65 AANAAATACATTGATATAGAAAACTAGAAAAAGTAAATGACATATTAATTA 124  
DB 1234 AATTTAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATA 1293  
QY 125 ACATGAAAAAGAAATTTATTAACAAGACTGAGAACGTTATTAATTGAATGATTA 184  
DB 1294 AAAAAATATATAATAATAATAATAATAATAATAATAATAATAATAATA 1353  
QY 185 AATTGAAAACTGCATCTGAAAGCAACTTTATTTGTAATTTCTTAATGATGCTT 244  
DB 1354 AAAAAATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1413  
QY 245 TTATGACTATATACATGATTTTTCATATAGAAACCATGTTAAAAATTTTATTTA 304  
DB 1414 TTTAAAGTTTGAATAATAATAATAATAATAATAATAATAATAATAATA 1470  
QY 305 AAAATTAAGCTGTGCTCAAGCTGATCATATTTCTTTATTTGATTTGGAGAAAT 364  
DB 1471 AAAATGAGCTTTTAAAGCCCAATTTTTCATATACGTAATATGACGTTCAAGTT 1530  
QY 365 ACTGTTCTGATACATGAAATGCAAAATTTTGATTTTAATCTCNCATATTTTAA 424  
DB 1531 TTTATGCTACTTCTTAACATTAAGTATTTCTTTATTTTAAAGCTTTTCTTAA 1590  
QY 425 ACTATGGAATGATTATGATGATGATGATGATGATGATGATGATGATGAT 472  
DB 1591 GCTTTATTTTCTTTTAATACATTTATTTCTTTTCTTTTCTTTTCTTT 1638

RESULT 11  
US-08-998-416-186/C  
Sequence 186, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Utegen  
APPLICANT: Knechtel, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwalls Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 7.5%; Score 51.2; DB 3; Length 615;  
Best Local Similarity 46.5%; Pred. No. 0.03;  
Matches 265; Conservative 0; Mismatches 297; Indels 8; Gaps 3;

QY 6 AATTTCTATTAACATTTCTTCAAGCATTTATTTATCTCATCTCTCAATTTTA 65  
DB 573 AATTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 514  
QY 66 ANAAATACATTGATATAGAAAACTAGAAAAAGTAAATGACATATTAATTA 125  
DB 513 TTAATATATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 454  
QY 126 CATGAAAAAGAAATTTATTAACAAGACTGAGAACGTTATTAATTGAATGATTA 185  
DB 453 CTATATAAAAGTTAAATTAATTAATCAACATATATTTTAAATAATGATATATA 394  
QY 186 ATTTGAAACTGC---ATCGAAAGCAACTTTATTTGTCATTTCTTAATGATGCTG 242  
DB 393 AAAATTAATTTTACATATTTTAAATTAATTAATTTTAAATTAATTAATTA 334  
QY 243 TTTATGACTATATAC--ACTGATTTTTCATATAGAAACCATGTTAAAAATTTTAT 300  
DB 333 TTAATTAACATTAATAATAATAATAATAATAATAATAATAATAATAATA 274  
QY 301 TTTAAAAATTAAGCTGTGCTCAAGCTGATCATATTTCTTTATTTTGAATTTGGA 360  
DB 273 AGAAATTAATTAATCTAATTAATTAATTTTAACTAATTTAAATTTGAACATAGA 214  
QY 361 AATATCTGTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 420  
DB 213 ATGATATTCATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTA 154  
QY 421 AAGAACTATGAGAAATGATTAATGATGATGATGATGATGATGATGATGAT 477  
DB 153 GTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 94  
QY 478 GCTGTTGCAATGTTGTTCTTACTAGTTCTCCAGAGAAACCTTAATCTGAATCTTC 537  
DB 93 ATGATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 34  
QY 538 AGCGAATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 567  
DB 33 ATCTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATTA 4

RESULT 12

US-09-601-198-63  
Sequence 63, Application US/09601198  
Patent No. 6531583  
GENERAL INFORMATION:  
APPLICANT: Casseil, Gail H.  
APPLICANT: Chen, Elison Y.  
APPLICANT: Glass, Jennifer S.  
APPLICANT: Glass, John I.  
APPLICANT: Heiner, Cheryl R.  
APPLICANT: Letkowitz, Elliot  
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA  
FILE REFERENCE: UAB-13452/22  
CURRENT APPLICATION NUMBER: US/09/601,198  
CURRENT FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/073,189  
PRIOR FILING DATE: 1998-01-30  
NUMBER OF SEQ ID NOS: 181  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 63  
LENGTH: 1692  
TYPE: DNA  
ORGANISM: Ureaplasma urealyticum  
US-09-601-198-63  
Query Match 7.4%; Score 51; DB 4; Length 1692;  
Best Local Similarity 47.0%; Pred. No. 0.036;  
Matches 156; Conservative 0; Mismatches 176; Indels 0; Gaps 0;  
QY 118 TAACTTCATGAGAAAGAAATTAATACAAAGAGCTGAGACGTTAATTAATGAAATG 177  
DB 189 TACAGTAAACAAATAATACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 248  
QY 178 AGATTATTAATTGAAAGCTGATCGAATGAAAGCAACTTATGTTCAATTTCTTAATGA 237  
DB 249 AGATTTCATTAATAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 308  
QY 238 TGGTCTTTTATGACTTAATACACTGATTTTTCATTAAGAAACCATGTTAAATATTTT 297  
DB 309 CTCTTATTTATCATTATTAAGTTTATTAAGCAATTAATTAATTAATTAATTAATTAAT 368  
QY 298 TATTTTAAATAAGCCGTGCTGATCAATTTCTTTATTTTATTTGATTTGGGA 357  
DB 369 AGATGATTTGTAGATTATTAATCAATCTTTTGAATGATTTTATTTATTTGATCAACA 428  
QY 358 AGAAATTAATCTGTTCTGATAGCATGAATGCAAAATTTTGAATTTTATCTGCTAAT 417  
DB 429 AGAAAGAAATAATGAAATACCATTAAGAAATTAATTAATTAATTAATTAATTAATTAAG 488  
QY 418 TTTAAGAACTATGGAATGATTAATGACA 449  
DB 489 TAAATATAGTATTCCTCAATTTAAATTAATCAAA 520  
RESULT 13  
US-08-213-419B-3/C  
Sequence 3, Application US/08213419B  
Patent No. 6333406  
GENERAL INFORMATION:  
APPLICANT: Inesburg, J. et al.  
TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM  
FILE REFERENCE: J11-002CNC  
CURRENT APPLICATION NUMBER: US/08/213,419B  
CURRENT FILING DATE: 1994-03-14  
PRIOR APPLICATION NUMBER: US 07/870,506  
PRIOR FILING DATE: 1992-04-17  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 6124  
TYPE: DNA  
ORGANISM: Plasmodium falciparum

FEATURE:  
NAME/KEY: CDS  
LOCATION: (2407)..(2439)  
NAME/KEY: CDS  
LOCATION: (2598)..(3404)  
NAME/KEY: CDS  
LOCATION: (3580)..(3720)  
NAME/KEY: CDS  
LOCATION: (3850)..(5835)  
US-08-213-419B-3  
Query Match 7.3%; Score 50.2; DB 4; Length 6124;  
Best Local Similarity 48.9%; Pred. No. 0.06;  
Matches 133; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
QY 77 TAGATTAGAAAACTGAAAAAGATTAATGAGATTAATTAATTAATTAATTAATTAATTAAT 136  
DB 6077 TATTTTATTAATAAAAAAAAAAAAAAAAAAAAAAAAAATTAATTAATTAATTAATTAAT 6018  
QY 137 AAATTAATTAAGAGCTGAGAACTTATTAATTAATTAATTAATTAATTAATTAATTAAT 196  
DB 6017 ATTAATTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAATTAATTAATTAATTAATTAAT 5958  
QY 197 GCATCTGAAGCAAACTTATTTGTTCAATTAATTAATTAATTAATTAATTAATTAATTAAT 256  
DB 5957 GATAAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5898  
QY 257 CACTGATTTTTCATTAAGGAAACCATGTTAAATAATTTTATTTTAAATAAGCTG 316  
DB 5897 TAAATAATTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5838  
QY 317 TGTCAAGCTCTGATCATATTTCTTTATTTT 348  
DB 5837 TATCATTAACAGAAATTAATTAATTTGT 5806  
RESULT 14  
US-08-998-416-1137/C  
Sequence 1137, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippesen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII  
FILE REFERENCE: J11-002CNC  
CURRENT APPLICATION NUMBER: US/08/998,416  
CURRENT FILING DATE: 1997-12-10  
PRIOR APPLICATION NUMBER: CH 0016/97  
PRIOR FILING DATE: 31-DEC-1996  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 36,241  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 36,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 1137:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: PAG1692RP  
US-08-998-416-1137

Query Match 7.2%; Score 49.2; DB 3; Length 636;  
Best Local Similarity 45.8%; Pred. No. 0.076;  
Matches 240; Conservative 0; Mismatches 277; Indels 7; Gaps 2;

47 TATATCTGACGAAATTTTAAANAATAACATTAGTAAAGAAAGCTAGAGAAAAGATNA 106  
523 TATATTTTATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTTTA 464  
107 ATGCAATTAATTAACCTTACATGAAAAAGAAATTAATTAACAAAGACTGAGAACGTTAT 166  
463 ATAAATTAATCTTATTAATAAGATTAAATTAATTAATTAATTAATTAATTAATGA 404  
167 AAATGAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 226  
403 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 344  
227 ATCTTAATGATGATGTTTAACTAATTAATTAATTAATTAATTAATTAATTAATTAAT 286  
343 ATATATTTTATTAATAACATTTATTAATAATAATAATAATAATAATAATAATAATAAT 288  
287 AAAAATTTTATTTTAAATAAAGCCGTGCTCAAGCTGATCATATTTCTTTTAT 346  
287 TAAATTAATTTTAAAGAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 228  
347 TTGATTTGGGAAGAAATACCTGTTCTGATGATGATGATGATGATGATGATGATGAT 406  
227 TGAACATGACTAAATGATATTCATTAATTAATTAATTAATTAATTAATTAATTAAT 168  
407 ATCTCCTAATTTTAAAGAACTTATTAATTAATTAATTAATTAATTAATTAATTAAT 465  
167 TAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 108  
466 -AATTAGCCAGCTGTTGCAATGTTGTTTCTTACTTACTTCTCCCAAGAAAGCT 523  
107 TAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 48  
524 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 567  
47 AAAAGTTTATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 4

## RESULT 15

US-08-487-826B-13  
Sequence 13; Application US/08487826B  
Patent No. 5993827

## GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chelnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knodbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/487,826B  
APPLICATION NUMBER: DATA  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelien, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-0176  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 7.1%; Score 48.6; DB 2; Length 19124;  
Best Local Similarity 45.7%; Pred. No. 0.14;  
Matches 200; Conservative 0; Mismatches 237; Indels 1; Gaps 1;

15 TTAACATTTCTTCAAGCAATTAATTTATCTATATCTGATGATTTTAAANAATAAC 74  
1757 TTTAGCACTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1816  
1817 ATTAATTAATCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1876  
135 GGAATTAATTAACAAAGACTGAGAACGTTATTAATTAATTAATTAATTAATTAAT 194  
1877 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1936  
195 CTGATCTGAAGCAAACTTATTTGTTCAATTAATTTTAAATTAATTTTAAATTAAT 254  
1937 ATGGAATGTTTGTATATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1995  
255 TACATGATTTTCAATTAAGAAAGCCGATGATGATGATGATGATGATGATGATGAT 314  
1996 TTTTATTTTATTTTGTGTTCTTAATTAATTAATTAATTAATTAATTAATTAAT 2055  
315 TGTGTTCAAGCTGATCAATTTCTTTATTTTGAATTTGGGAAGAAATACGTTCTG 374  
2056 ATATTAATTAATTAAGCAAAATTAATTTTATTAATTAATTAATTAATTAATTAAT 2115  
375 ATAGCATGAAGCAAAATTTTGAATTTTATTAATTTTAAATTTTAAATTAATTAAT 434  
2116 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2175  
435 AATTGATTAATGAATGA 452  
2176 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2193

Search completed: November 27, 2003, 12:32:17  
Job time: 51.7082 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:02:42 ; Search time 4026.04 Seconds

(without alignments) 10100.295 Million cell updates/sec

Title: US-09-835-992a-21

Perfect score: 994  
Sequence: 1 ctccaccgctgcctccctcag.....ccaccctgggctgctgggat 994Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_ry:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: gb\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_ets:\*  
28: em\_vl:\*  
29: em\_un:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pin:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_ry:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	978	98.4	994	6	AR146581 Sequence
2	978	98.4	994	6	BD079830 Cancer-as
3	765.8	77.0	2572	9	HUMSCP2A
4	646.8	65.1	2661	4	AF051897
5	576.8	58.0	1836	10	RATNSLTP
6	576.8	58.0	1945	10	RNNSLTPR
7	576.6	58.0	1980	10	BC034613
8	576.6	58.0	2626	10	BC018384
9	575.2	57.9	2571	10	RATSCPXA
10	575.2	57.9	2599	10	RAT60KDA
11	573.4	57.7	2152	6	AX306091
12	573.4	57.7	2152	10	MUSSCP
13	522.2	52.5	1409	10	RATSCP2
14	508.2	51.1	1500	9	HUMSCP2B
15	425.6	42.8	2443	5	CHKSCP2A
16	280.2	28.2	1359	3	DB6473
17	267.6	26.9	1916	3	AY089539
18	266	26.8	1375	3	AY058779
19	248.6	25.0	30177	3	CE57A10C
20	243	24.4	1376	3	DMSCPXRT
21	243	24.4	1804	3	DMSCPX
22	241.4	24.3	87634	2	CEK01F6
23	214	21.5	56414	1	MTV002
24	214	21.5	307550	1	BX248343
25	211.4	21.3	167777	3	AC009204
26	211.4	21.3	172061	3	AC009203
27	211.4	21.3	232737	3	AE003660
28	208.2	20.9	85699	2	AC020021
29	208.2	20.9	167777	3	AC009204
30	208.2	20.9	172061	3	AC009203
31	208.2	20.9	232737	3	AE003660
32	203	20.4	128469	2	AC020029
33	199.6	20.1	14995	1	AE007112
34	196.2	19.7	300425	1	AP005044
35	184.4	18.6	103284	8	NCB9B11
36	161.6	16.3	785	10	MUSSCP
37	154.4	15.5	201	9	HSCCPXT09
38	154.4	15.5	193774	9	AL445183
39	145.2	14.6	198	9	HSCCPXT10
40	132.6	13.3	40281	1	MLCB22
41	132.6	13.3	348450	1	MLEPRTN4
42	131.2	13.2	2381	1	AP306830
43	125.4	12.6	13287	1	AE011116
44	119.2	12.0	1725	9	HSCCPXT12
45	118	11.9	248677	2	AL844206

## ALIGNMENTS

RESULT 1  
AR146581  
LOCUS AR146581 994 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 21 from patent US 6218521.  
ACCESSION AR146581  
VERSION AR146581.1 GI:15109770  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 994)  
Obata, Y.  
Isolated nucleic acid molecules associated with gastric cancer and  
methods for diagnosing and treating gastric cancer  
JOURNAL Patent: US 6218521-A 21 17-APR-2001;

FEATURES Location/Qualifiers  
source 1..994  
/organism="unknown"  
BASE COUNT 313 a 168 c 249 g 248 t 16 others  
ORIGIN

Query Match 98.4%; Score 978; DB 6; Length 994;  
Best Local Similarity 100.0%; Pred. No. 4.2e-234;  
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCACCCAGTGTCTCTCAGATGTTGGGTATGCGAAAAAGAACATATGAAAAATATG 60  
DB 1 CTCACCCAGTGTCTCTCAGATGTTGGGTATGCGAAAAAGAACATATGAAAAATATG 60  
QY 61 GAACAAAAATGAACTTTTGGCAAAATTTGATGAAAAATCTAAACATTCACTTAATA 120  
DB 61 GAACAAAAATGAACTTTTGGCAAAATTTGATGAAAAATCTAAACATTCACTTAATA 120  
QY 121 ACCCGTATTCAGATTCGAAGATGATGATGATGATGATGATGATGATGATGATGATG 180  
DB 121 ACCCGTATTCAGATTCGAAGATGATGATGATGATGATGATGATGATGATGATGATG 180  
QY 181 TTTTGGATTTTGAATCTTACATGTTGTCCTCACTTCAGATGATGATGATGATGATG 240  
DB 181 TTTTGGATTTTGAATCTTACATGTTGTCCTCACTTCAGATGATGATGATGATGATG 240  
QY 241 TTTTGGCCGTAAGATTTTGAACAGATGATGATGATGATGATGATGATGATGATGATG 300  
DB 241 TTTTGGCCGTAAGATTTTGAACAGATGATGATGATGATGATGATGATGATGATGATG 300  
QY 301 TGGCACAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
DB 301 TGGCACAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
QY 361 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
DB 361 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
QY 421 CACCAAAATATATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
DB 421 CACCAAAATATATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
QY 481 CTTATGAAACATGGAAGCTCTGTCGAGAGACAAAGCTGATGATGATGATGATGATG 540  
DB 481 CTTATGAAACATGGAAGCTCTGTCGAGAGACAAAGCTGATGATGATGATGATGATG 540  
QY 541 ATATATCATATGAGAGAAAGTGGTCTAATCTTATGATGATGATGATGATGATGATG 600  
DB 541 ATATATCATATGAGAGAAAGTGGTCTAATCTTATGATGATGATGATGATGATGATG 600  
QY 601 ACCCACTAGCGCTAAGGCTCTGCTCAATGTCGAAAGCTGCTGCGAGCTGAGAGGGG 660  
DB 601 ACCCACTAGCGCTAAGGCTCTGCTCAATGTCGAAAGCTGCTGCGAGCTGAGAGGGG 660  
QY 661 AAGCCGGAAGAAAGGCAAGCTCTGTCGAAAGGCTGTCGNCATATTTTANGAT 720  
DB 661 AAGCCGGAAGAAAGGCAAGCTCTGTCGAAAGGCTGTCGNCATATTTTANGAT 720  
QY 721 TGGAGAACTGTGTTGTAACCTCTACAAGATGGGGTTTTCCCGAAGCCGCAAGTTCC 780  
DB 721 TGGAGAACTGTGTTGTAACCTCTACAAGATGGGGTTTTCCCGAAGCCGCAAGTTCC 780  
QY 781 TTTTGAACATCTCAAAATTTGAAGCCNGTTCCAAAGCTCTGCAAGTATNGTTTAA 840  
DB 781 TTTTGAACATCTCAAAATTTGAAGCCNGTTCCAAAGCTCTGCAAGTATNGTTTAA 840  
QY 841 NNNAAATCTGTTTAAAGNGATTTGAGAGAAACNTNAAAGGGAAGGGAAGAA 900  
DB 841 NNNAAATCTGTTTAAAGNGATTTGAGAGAAACNTNAAAGGGAAGGGAAGAA 900  
QY 901 ATTTGTGAAGAAATATGCGAGGAAATTTTGGCCCTTCAGGGGAAANAATGCGCTTGG 960  
DB 901 ATTTGTGAAGAAATATGCGAGGAAATTTTGGCCCTTCAGGGGAAANAATGCGCTTGG 960

QY 961 GGGGTAAAGANGGCCACCTGGGGTGGGGAT 994  
DB 961 GGGGTAAAGANGGCCACCTGGGGTGGGGAT 994

RESULT 2  
LOCUS BD079830 994 bp DNA linear PAT 27-AUG-2002  
DEFINITION Cancer-associated nucleic acids and polypeptides.  
ACCESSION BD079830  
VERSION BD079830.1 GI:22625433  
KEYWORDS JP 2001516009-A/496.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE Old, L.J., Scanlan, M.J., Stockert, E., Gure, A., Chen, Y.T., Gout, I.,  
Oghare, M., Obata, Y., Pfeundscher, M., Tureci, O. and Sahin, U.  
Cancer-associated nucleic acids and polypeptides  
Patent: JP 2001516009-A 496 25-SEP-2001;  
LUDWIG INSTITUTE FOR CANCER RESEARCH  
OS Homo sapiens (human)  
PN JP 2001516009-A/496  
PD 25-SEP-2001  
PF 15-JUL-1998 JP 2000503425  
PR 17-JUL-1997 US 08/896164, 10-OCT-1997 US 60/061599 PR  
10-OCT-1997 US 60/061765, 10-OCT-1997 US 08/948705 PR  
11-OCT-1997 GB 9721697.2, 22-JUN-1998 US 09/102322 PI  
J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI  
PI CHEN, IYAN GOUT, MICHAEL O'HARE, YUICHI OBATA, MICHAEL PFEUNDSCHUH, PI  
OZLEM TURECI,  
PI UGUR SAHIN  
PC G01N33/574, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61K48/00, PC  
A61P35/00,  
PC C07K14/82, C07K16/32, C12N15/09//C07K16/46, C12P21/08, A61K37/02,  
PC C12N15/00  
CC Cancer-associated nucleic acids and polypeptides. FH Key  
Location/Qualifiers  
FT source 1..994  
/organism="Homo sapiens (human)"  
location/Qualifiers  
source 1..994  
/organism="Homo sapiens"  
/mol type="genomic DNA"  
/db xref="taxon:9606"

BASE COUNT 313 a 168 c 249 g 248 t 16 others  
ORIGIN

Query Match 98.4%; Score 978; DB 6; Length 994;  
Best Local Similarity 100.0%; Pred. No. 4.2e-234;  
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCACCCAGTGTCTCTCAGATGTTGGGTATGCGAAAAAGAACATATGAAAAATATG 60  
DB 1 CTCACCCAGTGTCTCTCAGATGTTGGGTATGCGAAAAAGAACATATGAAAAATATG 60  
QY 61 GAACAAAAATGAACTTTTGGCAAAATTTGATGAAAAATCTAAACATTCACTTAATA 120  
DB 61 GAACAAAAATGAACTTTTGGCAAAATTTGATGAAAAATCTAAACATTCACTTAATA 120  
QY 121 ACCCGTATTCAGATTCGAAGATGATGATGATGATGATGATGATGATGATGATGATG 180  
DB 121 ACCCGTATTCAGATTCGAAGATGATGATGATGATGATGATGATGATGATGATGATG 180  
QY 181 TTTTGGATTTTGAATCTTACATGTTGTCCTCACTTCAGATGATGATGATGATGATG 240  
DB 181 TTTTGGATTTTGAATCTTACATGTTGTCCTCACTTCAGATGATGATGATGATGATG 240  
QY 241 TTTTGGCCGTAAGATTTTGAACAGATGATGATGATGATGATGATGATGATGATGATG 300



Db 241 TTTGGCCGTAAGATTTGTACAGAAATATGGCTGCATCCAAAGCTGGGAAATTT 300  
Qy 301 TGGCACAAGAAATGATGACTGATTTGGCCAAAGCTGTTTGAAGAAAAAGCATTTATTTAA 360  
Db 301 TGGCACAAGAAATGATGACTGATTTGGCCAAAGCTGTTTGAAGAAAAAGCATTTATTTAA 360  
Qy 361 TGGTGGCTTTGATTTAGTAAAGAAAGCTGCAAGAAATGCTATGAGAAATCTGGGCTGA 420  
Db 361 TGGTGGCTTTGATTTAGTAAAGAAAGCTGCAAGAAATGCTATGAGAAATCTGGGCTGA 420  
Qy 421 CACCAATGATATTTGACGTATAGAACTTCAAGATTTCTTTCTACCAAGAACTCTTA 480  
Db 421 CACCAATGATATTTGACGTATAGAACTTCAAGATTTCTTTCTACCAAGAACTCTTA 480  
Qy 481 CTTATGAAGCACTGGGACTCTGTCCAGAAAGCAAGGTGCAACGCTGGTTGATAGAGAG 540  
Db 481 CTTATGAAGCACTGGGACTCTGTCCAGAAAGCAAGGTGCAACGCTGGTTGATAGAGAG 540  
Qy 541 ATTAATACATATGAGAGAAAGTGGGTCTAATACTAGTGGTGGACTGATTTCAAGAGGAC 600  
Db 541 ATTAATACATATGAGAGAAAGTGGGTCTAATACTAGTGGTGGACTGATTTCAAGAGGAC 600  
Qy 601 ACCCACTAGGCGCTACAGGCTCTTGTCTCACTGTGCAAGAACTCTGCTGGCAGTGAAGGGG 660  
Db 601 ACCCACTAGGCGCTACAGGCTCTTGTCTCACTGTGCAAGAACTCTGCTGGCAGTGAAGGGG 660  
Qy 661 AAGCCGAAAAGAGCAAGAAAGTCTGCTGGTCAAGAAAGTGGCTCTGCGCATATTTTANGCAT 720  
Db 661 AAGCCGAAAAGAGCAAGAAAGTCTGCTGGTCAAGAAAGTGGCTCTGCGCATATTTTANGCAT 720  
Qy 721 TGGAGAACTGTGGTGTGTAACACTCAAGAAAGTGGGTCTTCCGGAAGCCGCAAGTTC 780  
Db 721 TGGAGAACTGTGGTGTGTAACACTCAAGAAAGTGGGTCTTCCGGAAGCCGCAAGTTC 780  
Qy 781 TTTTGAATCTCAAAATTTGAAGCCGTTTCCAAACCAAGCTCTGCAAGTATNGTTTAA 840  
Db 781 TTTTGAATCTCAAAATTTGAAGCCGTTTCCAAACCAAGCTCTGCAAGTATNGTTTAA 840  
Qy 841 NNNAAATCTGTTTAAAGGNGATTTGAAGAAAGAACTTNAAGGGAAGGGAAGACA 900  
Db 841 NNNAAATCTGTTTAAAGGNGATTTGAAGAAAGAACTTNAAGGGAAGGGAAGACA 900  
Qy 901 ATTGTGAAGAAAGAAATTCGNGGGAATTTTGGCTTGAAGGGAAGAAATGCGCCCTGG 960  
Db 901 ATTGTGAAGAAAGAAATTCGNGGGAATTTTGGCTTGAAGGGAAGAAATGCGCCCTGG 960  
Qy 961 GGGGTAAAGAGAGCCCAAGCTGGGTGGTGGAT 994  
Db 961 GGGGTAAAGAGAGCCCAAGCTGGGTGGTGGAT 994

RESULT 3  
HUMSCP2A 2572 bp mRNA linear PRI 06-DEC-1993  
LOCUS Human sterol carrier protein X/sterol carrier protein 2 mRNA,  
DEFINITION complete cds.  
ACCESSION M75883  
VERSION M75883.1 GI:432974  
KEYWORDS sterol carrier protein-2, sterol carrier protein X.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
REFERENCE 1 (bases 1 to 2572)  
AUTHORS He Z., Yamamoto R., Furch B.E., Schantz L.J., Naylor S.L.,  
George H., Billheimer J.T. and Strausz J.F. III.  
CDNA encoding members of a family of proteins related to human  
sterol carrier protein 2 and assignment of the gene to human  
chromosome 1 p21----pter  
JOURNAL DNA Cell Biol. 10 (8), 559-569 (1991)  
MEDLINE 92029618  
PUBMED 1718316

REFERENCE 2 (bases 1 to 2572)  
AUTHORS Veen J., Helsten E., Branoski B.L., Emanuel B.S., Billheimer J.T.,  
Mead S., Cowell J.K., Strausz J.F. III. and Pelcon L.  
TITLE Assignment of sterol carrier protein X/sterol carrier protein 2 to  
1p32 and exclusion as the causative gene for infantile neuronal  
ceroid lipofuscinosis  
JOURNAL Unpublished  
COMMENT On Dec 6, 1993 this sequence version replaced gi:410029.  
ORIGINAL source text: Human liver cDNA to mRNA.  
FEATURES  
Source location/Qualifiers  
1..2572  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="1p32"  
/tissue\_type="liver"  
/clone\_11b="Clontech HL-1001b;HL-1115b"  
1..2572  
/gene="SCP-X/SCP-2"  
22..1665  
/gene="SCP-X/SCP-2"  
/codon\_start=1  
/product="sterol carrier protein-X/sterol carrier  
protein-2"  
/protein\_id="AAA03557.1"  
/db\_xref="GI:432975"  
/translation="MSSSPMRPATIRRVFVGVGWTFRKRGAEASRDYPDIAERAG  
KALDAQIPYSAVDACVGVFGDSTGQRAIYSLGNTGIPINNNCATGSLALF  
MAROLIQGVABECVLALEPKMSKSLGDKSDRTIPNDKVDLLINKYLSAHVAP  
OMFGVAGKEHEKYSTKIEHPAKISGMKHKHSVNNPQPOPEYLDDEVMAKEVDF  
LTILOCCPTSGAAAAILASEAFVQYGLQSAVAEIILOEMMTDLPSSFEESIIKMY  
GPDMSKEARRCYERKSGLTDPNDIVIELHDCPESTNELITVETALGICPGGQATLVDRG  
DNTYGRKVINPSGSLISKHPLATGTAOCCEBLCQLRGEGKQVPRKALALNHL  
GIGAVVTLTKMGPEAASSFRTHQIAVPTSSASDQKALVKEIKKUEBGO  
FVKIGGIFARKVXDPGKGAATWVDVNGKSVLPNSDKAADCTITMADSDFLALM  
TGKMPQSAFEGCKLKLTGNMGLAMKLNLOLPENATL"  
1932..1937  
/gene="SCP-X/SCP-2"  
/evidence="not\_experimental"  
2213..2218  
/gene="SCP-X/SCP-2"  
/evidence="not\_experimental"  
2258..2263  
/gene="SCP-X/SCP-2"  
/evidence="not\_experimental"  
2341..2346  
/gene="SCP-X/SCP-2"  
/evidence="not\_experimental"  
2550..2555  
/gene="SCP-X/SCP-2"  
/evidence="experimental"  
2572  
/gene="SCP-X/SCP-2"  
/evidence="experimental"  
BASE COUNT 794 a 439 c 570 g 769 t  
ORIGIN  
Query Match 77.0%; Score 765.8; DB 9; Length 2572;  
Best Local Similarity 91.1%; Pred. No. 5.9e-181;  
Matches 897; Conservative 0; Mismatches 68; Indels 20; Gaps 8;  
Qy 1 CTCACCCAGTGGCTCTCAGATGTTGGGTATGCTGGAAGAAACATATGAAAAATATG 60  
Db 485 CTCACCCAGTGGCTCTCAGATGTTGGGTATGCTGGAAGAAACATATGAAAAATATG 544  
Qy 61 GAACAAAATTAACACTTTGGCAAAATTTGATGAGAAATATCATTAACATTCAGTTAATA 120  
Db 545 GAACAAAATTAACACTTTGGCAAAATTTGATGAGAAATATCATTAACATTCAGTTAATA 604  
Qy 121 ACCCGTATTCAGTTCAGAGTGAATACATTTAGATGAAGTGAATGATGATTAAGAAG 180  
Db 605 ACCCGTATTCAGTTCAGAGTGAATACATTTAGATGAAGTGAATGATGATTAAGAAG 664

```

QY 181 TTTTGGATTTTGGACTATCTTACAAATGTGTCCCACTCAGATGGTGTGGAGAGCAA 240
Db 665 TTTTGGATTTTGGACTATCTTACAAATGTGTCCCACTCAGATGGTGTGGAGAGCAA 724
QY 241 TTTTGGCCAGTGAAGCATTGTGACAGAAATGAGCTGTGCAATCCAAAGCTGTGAATTT 300
Db 725 TTTTGGCCAGTGAAGCATTGTGACAGAAATGAGCTGTGCAATCCAAAGCTGTGAATTT 784
QY 301 TGGCACAAGAAATGATGATGATTTTCCCAAGCTGTTTGAAGAAAAAGCATTTATTTAAA 360
Db 785 TGGCACAAGAAATGATGATGATTTTCCCAAGCTGTTTGAAGAAAAAGCATTTATTTAAA 844
QY 361 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 845 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904
QY 421 CACCAATGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 905 CACCAATGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 964
QY 481 CTTATGAGCACTGGGACTCTGTCTCCAGAAAGACAGAGTCCAAAGCTGTGATAGAGAG 540
Db 965 CTTATGAGCACTGGGACTCTGTCTCCAGAAAGACAGAGTCCAAAGCTGTGATAGAGAG 1024
QY 541 ATTAATCATATGAGGAGAAAGTGGGTCAATCTTATGATGATGATGATGATGATGATGAT 600
Db 1025 ATTAATCATATGAGGAGAAAGTGGGTCAATCTTATGATGATGATGATGATGATGATGAT 1084
QY 601 ACCCACTAGAGCTGACAGGCTCTGTCTGATGATGATGATGATGATGATGATGATGATGAT 660
Db 1085 ACCCACTAGAGCTGACAGGCTCTGTCTGATGATGATGATGATGATGATGATGATGATGAT 1144
QY 661 AAGCCGGAAGAAAGCAAGGCTCTGTCTGATGATGATGATGATGATGATGATGATGATGAT 720
Db 1145 AAGCCGGAAGAAAGCAAGGCTCTGTCTGATGATGATGATGATGATGATGATGATGATGAT 1202
QY 721 TGGAGAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 1203 TGGAGAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1259
QY 781 TTTTGAACCTCATCAAAATTTGAAGCCGCTTCCAAAGCTGTGCAAGTATGATTTTAA 840
Db 1260 TTTTGAACCTCATCAAAATTTGAAGCCGCTTCCAAAGCTGTGCAAGTATGATTTTAA 1316
QY 841 NNNAAAATCTGTTTAAAGAGGATTTGAAGAGAAACNTNAAAGAGGAGAGAGGAGAGAG 900
Db 1317 GCGCAATCTTGTGTTTAA--GGAGATTTGAAGAGAACTTGA---AGAGAGAGGAGAGAG 1370
QY 901 ATTTTGAAGAGAAATTTGCGGAGAAATTTTGCCTTCAAGGAGAGAAATGAGCCCTGG 960
Db 1371 GTTTTGAAGAGAAATCGGTGT-----ATTTTGTCTTCAAGGTGAGAG-ATGGCCCTGG 1424
QY 961 GGGGTAAAGAGAGCCAGCCCTGGGG 985
Db 1425 GGGTAAAGAGAGCCAGCCCTGGGTG 1449

```

```

RESULT 4
AF051897 2661 bp mRNA linear MAM 30-APR-1998
LOCUS AF051897
DEFINITION Oryctolagus cuniculus sterol carrier protein X (SCP2) mRNA,
complete cds.
ACCESSION AF051897.1 GI:3093861
VERSION AF051897
KEYWORDS Oryctolagus cuniculus (rabbit)
SOURCE Oryctolagus cuniculus
ORGANISM Oryctolagus cuniculus
REFERENCE 1 (bases 1 to 2661)
AUTHORS Weber,F.E., Werder,M., Boffelli,D. and Hauser,H.
TITLE In pre-sterol carrier protein 2 (SCP2) in solution the leader
peptide 1-20 is flexibly disordered and the residues 21-143 adopt

```

```

JOURNAL
REFERENCE 2 (bases 1 to 2661)
AUTHORS Weber,F.E., Werder,M., Boffelli,D. and Hauser,H.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1998) Kieferchirurgie, Universitaetspital,
Raemster, 100, Zurich 8091, Switzerland
FEATURES
source location/Qualifiers
1..2661
/organism="Oryctolagus cuniculus"
/mol_type="mRNA"
/db_xref="taxon:9986"
/cell_type="enterocytes"
1..2661
/gene="SCP2"
53..1696
/gene="SCP2"
/note="peroxisomal 3-oxoacyl coenzyme A thiolase"
/product="sterol carrier protein X"
/protein_id="AAC15422.1"
/db_xref="GI:3093802"
/translation="MSSSRKRLAPLPRVVGVTGKTFKPKETEDARDPYMAKAGQ
KALADQIPYAVEQACIGYVDSTCGRAVYHSLGLTPIINNNNCSGTALF
MGROLIOGMAECVLAIFEMERGSIGAKPPDRNPMDKLIDVLINXGSAHPAP
OMPSAGKEHMEKXTKEHRAKIGMKHKSUNNPYQPOKEYSLDPMVSRIPEF
LIVLQCCRTSDGAAAIILASEEFPYKRGLOSKAVELIAQENVTFPPSPFEKSIITKV
GPDMSKEAARCYEKSGLRPSDIDVIELHDLFSANELLTYALGCEPKGKALVDKG
DNVYQKWINIPSGGLISKHPLATATGACAEICWOLRGAKROYGAVALQHL
GLGGAAYVTLKMGFPEAASSFRTHQIEAASGDFKANLVKEIEKLEEGEQ
FVKIKGIFAKRVKDGPGKREATVWDVKNKGSVLPSPDKADCTIITADSLALM
TKMNPQSAFPGKIKITGNMGLMKLQNLQDQKXKL"
BASE COUNT 806 a 490 c 623 g 742 t
ORIGIN
Query Match 65.1%; Score 646.8; DB 4; Length 2661;
Best Local Similarity 88.0%; Pred. No. 3.8e-151;
Matches 757; Conservative 0; Mismatches 95; Indels 8; Gaps 5;
1 CTCACCAAGTGTCTCTCAGATGTTGGGTATGCTGAAAAAGACATATGAAAAATATG 60
Db 516 CCCACCAAGTGTCTCTCAGATGTTGGGTCTGCTGAAAAAGACATGAAAAATATG 575
QY 61 GAACAAAATTTGAACACTTTCAAAAATTTGATGAGAAATATCATTAAGATTCAAGTATA 120
Db 576 GAACAAAATTTGAACACTTTCAAAAATTTGATGAGAAATATCATTAAGATTCAAGTATA 635
QY 121 ACCGATATTCACAGTCCCAAGATGATACAGTTTATGATGATGATGATGATGATGATGATGAT 180
Db 636 ACCGATATTCACAGTCCCAAGATGATACAGTTTATGATGATGATGATGATGATGATGATGAT 695
QY 181 TTTTGAATTTTGAATCTTACATGATGTTGCCACTCTTCAAGATGATGATGATGATGATGATGAT 240
Db 696 TTTTGAATTTTGAATCTTACATGATGTTGCCACTCTTCAAGATGATGATGATGATGATGATGAT 755
QY 241 TTTTGGCCAGAGAACATTTTGTACAGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 756 TTTTGGCCAGAGAACATTTTGTACAGATGATGATGATGATGATGATGATGATGATGATGAT 815
QY 301 TGGCACAAGAAATGATGATGATTTTCCCAAGCTGTTTGAAGAAAAAGCATTTATTTAAA 360
Db 816 TGGCACAAGAAATGATGATGATTTTCCCAAGCTGTTTGAAGAAAAAGCATTTATTTAAA 875
QY 361 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 876 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 935
QY 421 CACCAATGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 936 CACCAATGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995
QY 481 CTTATGAGCACTGGGACTCTGTCTCCAGAAAGACAGAGTCCAAAGCTGTGATAGAGAG 540

```

the same globular fold as in mature SCP2

Db 996 CTTATGAGCACTAGACTCTGTCCGAGCAAGGAGGAGCACTGTTGACAGAGAG 1055  
Qy 541 AATAATCAATATGAGAGAAAGTGGTCAATAATCCATAGTGTGAGACTGATTTCAAGGAGC 600  
Db 1056 AATAATCACTACGAGAGAAAGTGGTCAATAATCCATAGTGTGAGATTTCAAGGAGC 1115  
Qy 601 ACCCACTAGAGCCCTCAGAGTCTTGCTCAGTGTGAGAACTGTGTGAGAGAGAGG 660  
Db 1116 ACCCCCTGGAGCTACAGGTCTGCTCAATGTGCCAATCTGTGTGAGAGAGAGG 1175  
Qy 661 AAGCCGAAAAGAGGCAAAAGTTCCTGTGCAAAAGTGGTCTGTGCAATTAATTANGCAT 720  
Db 1176 AAGCCGAAA--GAGGCAAGTTCCTGTGCAAAAGTGGTCTGTGCAATTAATTGGGCT 1233  
Qy 721 TGGAGAACTGTGTGTTGTAACACTCTACAAAGATGGGTTTCCCGAAGCCCACTTCC 780  
Db 1234 CGAGAGAGTGGGTGTCACACTACAAAGATGGG--TTTTCCTGAGAGCTCCAGTT-C 1290  
Qy 781 TTTTGAACACTCAAAATTTGAAGCCNGTTCACCAAGCTGTGCAAGTATNGGTTTA 840  
Db 1291 TTTCAAGAACCACTCAAA--TTGAGGCTGTCCAAAC--AGCTGTGAGAGTATGATTTA 1347  
Qy 841 NGMAAATCTNGTTTAAAG 860  
Db 1348 GGCAATCTTGTCTTTAAG 1367

RESULT 5  
RATSLTP 1836 bp mRNA linear ROD 27-APR-1993  
LOCUS Rat non-specific lipid transfer protein (nsl-TP) mRNA, 3' end.  
DEFINITION M58287.1 GI:205769  
ACCESSION M58287  
VERSION M58287.1 GI:205769  
KEYWORDS non-specific lipid transfer protein.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 1836)  
AUTHORS Owendorp, B.C., van Heuuden, G.P. and Wirtz, K.W.  
TITLE The amino acid sequence of rat liver non-specific lipid transfer  
protein (sterol carrier protein 2) is present in a high molecular  
weight protein: evidence from cDNA analysis  
JOURNAL Biochem. Biophys. Res. Commun. 168 (2), 631-636 (1990)  
MEDLINE 90241231  
PUBMED 2334427  
COMMENT Original source text: Rat liver, cDNA to mRNA.  
FEATURES  
source location/Qualifiers  
1..1836  
/organism="Rattus sp."  
/mol\_type="mRNA"  
/db\_xref="taxon:10118"  
/cbase\_type="liver"  
1..1836  
/gene="nsl-TP"  
<1..1548  
/gene="nsl-TP"  
/codon\_start=1  
/product="non-specific lipid transfer protein"  
/protein\_id="AAA41726.1"  
/db\_xref="GI:205770"  
/translation="RDYDPLAKEAQKALADRLQIPYSAVEACVGVYVGESTCGRAI  
YHSLGLTGIPIINANNSTGTALFMAQOLVQGLANCVLAGEKKEKSLGTRYS  
DRSNLEKHIIVLILNKYGSACPFAPOLGSAKHEMYGTGHPAKIGWKNHKS  
VNNPSODEVSLDEIKMSRPVDFLTVLOCCPTSDGAAAIIVSEEVQGHQSK  
AVEIVAOMVTDMSTPEKRSYIKMVGDMSEAKRKCEKSGLGSPVDVTEIADCF  
STNELITTEALGLCEPGCGAGAVDRGNTYGGKWTINBGLISGHPLEGATLQCA  
ELCQULRGAGKQVPGACVALQHNLDGAGAAVTLVYRGPFAASFFTHQISAPT  
SSAGDFKANLIFKEIEKLEEGEPVKIGIFAFKXKDPGGEKATWVADVNGK  
GSVLPDSDKKADCTITMADSDLLALMTGKMNQSAFQOKLXIAGNMGLAMTLOSLQ  
LQPKAKL."

BASE COUNT

500 a 392 c 503 g 441 t

## ORIGIN

Query Match 58.0%; Score 576.8; DB 10; Length 1816;  
Best Local Similarity 78.7%; Pred. No. 1.3e-133;  
Matches 759; Conservative 0; Mismatches 187; Indels 19; Gaps 6;

Qy 6 CCAGTGTGCTCTCAGATGTTGGGTATCTGGAAAAAGAAATATGAAAAATATGAAACA 65  
Db 373 CCATTGTGCTCTCAGACTGTTGGAGCCGTGGAAAAAGAAATATGAAAAATATGAAACA 432  
Qy 66 AAAAAATGAAACATTTGCAAAAAATGGATGAAAAATCATTAACATTCAATTATAACCCG 125  
Db 433 AAGGTGAACACTTTGCAAAAAATGGATGAAAAATCATTAACACTCAATTATAACCCG 492  
Qy 126 TATTTCCAGTTCCAGATGAATACGATTATGATGAAGTATGAGCATCTTAAGAAATTTT 185  
Db 493 TATTTCCAGTTCCAGATGAATACGATTATGATGAAGTATGAGCATCTTAAGAAATTTT 552  
Qy 186 GATTTTGGACTATCTTCAATGTTGTCCCACTTCAATGAGTGTGACAGCAAAATTTTG 245  
Db 553 GATTTTGGACTATCTTCAATGTTGTCCCACTTCAATGAGTGTGACAGCAAAATTTTG 612  
Qy 246 GCCAGTGAAGCATTTGTACAGAAATATGAGTATGAGTATGAGTATGAGTATGAGTAT 305  
Db 613 TCTAGTGAAGCATTTGTACAGAAATATGAGTATGAGTATGAGTATGAGTATGAGTAT 672  
Qy 306 CAGAAATGATGATGATGATTTGCAAGCTGCTTTGAAGAAAAAGCAATTAATTAATGCTT 365  
Db 673 CAGGATGATGATGATGATGATTTGCAAGCTGCTTTGAAGAAAAAGCAATTAATTAATGCTT 732  
Qy 366 GCGTTGATGATGATGATGATGATTTGCAAGCTGCTTTGAAGAAAAAGCAATTAATTAATGCTT 425  
Db 733 GCGTTGATGATGATGATGATTTGCAAGCTGCTTTGAAGAAAAAGCAATTAATTAATGCTT 792  
Qy 426 AATGATATGATGATGATGATTTGCAAGCTGCTTTGAAGAAAAAGCAATTAATTAATGCTT 485  
Db 793 AGTATGATGATGATGATGATTTGCAAGCTGCTTTGAAGAAAAAGCAATTAATTAATGCTT 852  
Qy 486 GAGACACTGGAGCTGTGTCAGAGAGACAAAGTGTCAAGCTGTGTTGATGAGAGATTAAT 545  
Db 853 GAGACACTGGAGCTGTGTCAGAGAGACAAAGTGTCAAGCTGTGTTGATGAGAGATTAAT 912  
Qy 546 ACATATGAGAGAAAGTGTGATTAATCTGATGAGTGTGATGATGATGATGATGATGATGAT 605  
Db 913 ACTTACGAGAGAAAGTGTGATTAATCTGATGAGTGTGATGATGATGATGATGATGATGAT 972  
Qy 606 CTAGCGCTTACAGCTCTGTCTCAGTGTGACAGAACTCTGTCTGACAGTGTGAGAGAGAGCC 665  
Db 973 CTGSGTGTCAAGCTGTGCTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1032  
Qy 666 GGAAG 725  
Db 1033 GGAAG 1090  
Qy 726 GAACTGTGTGTGTAACACTTCAAGATGGGTTTCCCGAAGCCGACAGTCTCTTTTA 785  
Db 1091 GAGCTGTGTGTGTAACACTTCAAGATGGGTTTCCCGAAGCCGACAGTCTCTTTTA 1147  
Qy 786 GAACTCATCAAAATTTGAAGCCGTTCCAAACAGCTGTGCAAGTATGATGATGATGATGAT 845  
Db 1148 GAACTCATCAAAATTTGAAGCCGTTCCAAACAGCTGTGCAAGTATGATGATGATGATGAT 1204  
Qy 846 AATCTGTTTAAAGAGAGATTTGAAGAGAAACNTAAAGAGAGAGAGAGAGAGAGAGAGAG 905  
Db 1205 AATCTGTTTAAAGAGAGATTTGAAGAGAAACNTAAAGAGAGAGAGAGAGAGAGAGAGAG 1258  
Qy 906 TGAAG 965  
Db 1259 TGAAG 1313  
Qy 966 AAAAG 970  
Db 1314 AGAAG 1318

RESULT 6  
RNNSLIPTR 1945 bp mRNA linear ROD 17-FEB-1997  
LOCUS Rat mRNA for 58 kDa protein containing the non-specific lipid  
DEFINITION transfer protein (sterol carrier protein 2).  
X60654.1 GI:56771  
VERSION non-specific lipid transfer protein; sterol carrier protein-2.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1  
AUTHORS Billheimer, J.T., Strehl, L.L., Davis, G.L., Straus, J.F. III and  
Davis, L.G.  
TITLE Characterization of a cDNA encoding rat sterol carrier protein2  
JOURNAL DNA Cell Biol. 9 (3), 159-165 (1990)  
MEDLINE 90253610  
PUBMED 2340090  
REFERENCE 2  
AUTHORS Seedorf, U. and Asseman, G.  
TITLE Cloning, expression, and nucleotide sequence of rat liver sterol  
carrier protein 2 cDNAs  
JOURNAL U. Biol. Chem. 266 (1), 630-636 (1991)  
MEDLINE 91093192  
PUBMED 1985920  
REFERENCE 3  
AUTHORS Ossendrop, B.C., Van Heusden, G.P., De Beer, A.L., Bos, K.,  
Schouten, G.L. and Wirtz, K.W.  
TITLE Identification of the cDNA clone which encodes the 58-kDa protein  
containing the amino acid sequence of rat liver non-specific  
lipid-transfer protein (sterol-carrier protein 2). Homology with  
rat peroxisomal and mitochondrial 3-oxoacyl-CoA thiolases  
JOURNAL Eur. J. Biochem. 201 (1), 233-239 (1991)  
MEDLINE 92007881  
PUBMED 1915369  
REFERENCE 4 (bases 1 to 1945)  
AUTHORS Ossendrop, B.C.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1991) B.C. Ossendrop, Centre for Biomembranes and  
lipid, Enzymology, State, University of Utrecht, Post Box 80054,  
3508 TB Utrecht, THE NETHERLANDS  
See also M58287.  
COMMENT Location/Qualifiers  
FEATURES  
source 1..1945  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-dawley"  
/db\_xref="taxon:10116"  
/sex="Female"  
/tissue\_type="liver"  
/clone\_lib="lambda gt11"  
/dev\_stage="adult"  
1..1930  
/evidence=experimental  
<1..1642  
/note="open reading frame for 58 kDa protein containing  
the non-specific lipid transfer protein"  
/codon\_start=2  
/protein\_id="CAA43061.1"  
/db\_xref="GI:56773"  
/translation="PSVALNSPLRVRVGVGWMTRKMKPGGNSRDYDPLAKEAGOK  
ALADROIYSAYSAVGAQGVVGSSTCGPRAIYHSILGICPIIINNKNSTSTALFM  
AQLVQGGILANCVLALGPRKMGSLGTYSDRNPLEHIDVLINNKMGSCPPAPQ  
LFGAGKEHMETYGRKVEHFAKIGMKNHSHVNNPSPQODEYSUDEIMKSRPVDPL  
TVLQCCPTSDGAAAIIVSEEFVQKHGQSKAVEIADQEMVTDMSTFEKSVIMVG  
YDMSKEAARKCYEKSGIGPSDVAIEIHDQFSTNELITYEALGLCEGGGALVVRG  
NYGGMVIVNPSGGLISKGPATGATGLACALCWLQREAGAKROVPAGKVALQHNIG  
LGGAIVTLIYRMGPBAASFRTHQISAAPTSADGPRANLIIFEIEKLEEBEERF

conflict  
33..34  
/citation=[2]  
/replace="c9"  
792  
/citation=[2]  
/replace="c"  
804..809  
/citation=[1]  
/replace="aaaaa"  
1211..1642  
/codon\_start=1  
/product="non-specific lipid transfer protein"  
/protein\_id="CAA43060.1"  
/db\_xref="GI:56772"  
/translation="MGFPEAASSPFRTHQISAAPTSADGPRANLIIFEIEKLEEBE  
EFVYKIGIFAFKVDGPGKREATWVDVKNKGSVLPDSDKKADCTITMADSDLLA  
LMTGMNPOSAPFQOKLKIAGNMGILAMTQSLQIDPDAKL"  
CDS  
mat\_peptide  
misc\_feature  
BASE COUNT 531 a 415 c 535 g 464 t  
ORIGIN  
Query Match 58.0%; Score 576.8; DB 10; Length 1945;  
Best Local Similarity 78.7%; Pred. No. 1.3e-133;  
Matches 759; Conservative 0; Mismatches 187; Indels 19; Gaps 6;  
QY 6 CCAAGTTCCTCCCAATGTTGGTATCTGGAAAAGAACATATGAAAAAATATGAAACA 65  
DB 467 CCATTGGCTCCACACTGTTGGAGCGCTGGAAAAGAACATATGAAACATATGAAACA 526  
QY 66 AAAATTGAACACTTGGCAAAAATTTGATGAAAAAATCATTAACATTGATTAACCG 125  
DB 527 AAGGTTGAACACTTGGCAAAAATTTGATGAAAAAATCATTAACACTGATTAACCG 586  
QY 126 TATTCACAGTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185  
DB 587 TATTCACAGTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 646  
QY 186 GATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 245  
DB 647 GATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 706  
QY 246 GCCAGTGAAGCATTTGTGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 305  
DB 707 TCTAGTGAAGCATTTGTGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 766  
QY 306 CAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 365  
DB 767 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 826  
QY 366 GGCCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 425  
DB 827 GGCCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 886  
QY 426 AATGATTTGAAGCATTTGTGACAGAGATGATGATGATGATGATGATGATGATGATGAT 485  
DB 887 AGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 946  
QY 486 GAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 545  
DB 947 GAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006  
QY 546 ACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605  
DB 1007 ACTTACGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1066  
QY 606 CTAGCGGTAAAGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 665

```

Db      1067 CTGGGTGCACAGCTGTGGCTCAGTGCAGCGGAGAGCTCTGTGGCAGCTGAGAGCGAAGCC 1126
Qy      666 GGAAGAAAGAGCAAGTTCTGTGTGCAAGAGTGGCTCTGTGNCATTAATTANCAATTGGAG 725
Db      1127 GGAAGAA--GAGGCAAGGTTCTGTGGGCAAGAGTGGCTCTGTGACAGCAATTAGGCTTTGGAG 1184
Qy      726 GAACGTGGTGTGTACACTCTACAGATGGGGTTTTCCGGAAGCGCCAGTTCTTTTA 785
Db      1185 GAGCTGCTGTGTCTACCTCTACAGATGGG--TTTTCCGAAAGCTGCCAGCTCC--TTCA 1241
Qy      786 GAACCTCATCAAAATTTGAAGCCNGTTCCACCAAGCTCTGCAATNATNGTTTAAAGNAA 845
Db      1242 GAAGCAGCAGATTTCA--GCTGCTCCACAGCTCTGACAGGGATGATTCAGAGCA 1298
Qy      846 AATCTNGTTTAAAGNGGATTTGAGAGAAACNTNAAAGAGGGAAGGGGAACATTTG 905
Db      1299 ATCTCATTTTAAAGAA-----ATCGAAGAAAGCTTGAAAGAGGAGGGAAGGTTG 1352
Qy      906 TGAAGAAATAATNCGANGGGAATTTTGGCCCTTCAAGGGGAATNATGGCCCTGGGGGT 965
Db      1353 TGAAGAAATTCGCTGCGC-----ATTGTCCTTCAAGTGAAGATGCGCCCGGGGCA 1407
Qy      966 AAAG 970
Db      1408 AGAG 1412

RESULT 7
LOCUS   BC034613
DEFINITION Mus musculus, steroid carrier protein 2, liver, clone MGC:29961
IMAGE:5123611, mRNA, complete cds.
ACCESSION BC034613
VERSION   BC034613.1 GI:21961462
KEYWORDS  MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  AUTHORS  Strausberg, R.
  TITLE    Direct Submission
  JOURNAL  Submitted (24-JUL-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK   NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT   Contact: MGC help desk
          Email: gcgphs-remail.nih.gov
          Tissue Procurement: Jeffrey E. Green, M.D.
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: Baylor College of Medicine Human Genome
          Sequencing Center
          Center code: BCM-HGSC
          Web site: http://www.hgsc.bcm.tmc.edu/cdna/
          Contact: amg@bcm.tmc.edu
          Guarnarone, P.H., Garcia, A.M., Lu, X., Huilyk, S.W., Hale, S.M.,
          Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
          Richards, S., Gibbs, R.A.

FEATURES
  SOURCE
    1..1980
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /map="FVB/N"
  Location/Qualifiers
    Clone distribution: MGC clone distribution information can be found
    through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
    Series: IBK plate: 42 Row: n Column: 7
    This clone was selected for full length sequencing because it
    passed the following selection criteria: Similarity but not
    identity to protein.
  
```

```

CDS
  1..1665
    /codon_start=1
    /product="steroid carrier protein 2, liver"
    /protein_id="AAH34613.1"
    /db_xref="GI:21961463"
    /translation="MPSVALKSPRLRVRVVGVTGKFMKPGSGNSRDYPMKAKEAQ
    KALSDAOIPIYSAVDAQCYGVYGDSTSGRAIYHSILGTGIPILVNNNSCTGTALE
    GYDMSKEARACYEKSGILTPNDVDVIELHDFSAVEIVAEQMTDLPSTFEKSIILKV
    DNTYGKAVINPISGGLISKGPLGATGACALCMQWRGKQVAGKVALQHL
    GLGGAVVTLVRMGPPEAASFRTHQVAAIPSSAGDFKANLVFKETIEKLEEGED
    FVKIKGTFARKVDGPGREKATVVDVYKNGKSVLPNSDKKADCTIMADSDLLALM
    TGNMPQSAFQGGKLIKTAGNGLAKLONLQPEKATL"
  BASE COUNT
    556 a 417 c 537 g 470 t
  ORIGIN
    1 CTCACCCAGTGTCTCTCAGATGTTGGGTATGCTGAGAAAGAAATATG 60
    Db 485 CACATCCGATTAATCTCTCAGATGTTGGGTATGCTGAGAAAGAAATATG 544
    Qy 61 GAACAAAATTTGAACATTTTGCAAAAATTTGATGAGAAATATCAATACTAGTTA 120
    Db 545 GAACAAAATTTGAACATTTTGCAAAAATTTGATGAGAAATATCAATACTAGTTA 604
    Qy 121 ACCGATATCCAGTCCAGATGATGATGATGATGATGATGATGATGATG 180
    Db 605 ACAGTATTTCCAGTCCAGATGATGATGATGATGATGATGATGATGATG 664
    Qy 181 TTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 240
    Db 665 TTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 724
    Qy 241 TTTTGGCCAGTGAACATTTTGAATTTTGAATTTTGAATTTTGAATTT 300
    Db 725 TTCTGTCCAGGAGAGATTTTGTGTCAGAGTACGCGCTCAGACCCAAAGGGGAGATTT 784
    Qy 301 TGGCAAGAAATGATGATGATTTTGGCAAGCTCTTGAAGAAAGCAATTATAA 360
    Db 785 TGGCCAGAGATGATGATGATTTTGGCAAGCTCTTGAAGAAAGCAATTATAA 844
    Qy 361 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATG 420
    Db 845 TGGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATG 904
    Qy 421 CACCAATGATATGATGATGATGATGATGATGATGATGATGATGATG 480
    Db 905 CACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 964
    Qy 481 CTTATGAAGCACTGGGATCTGTCTCAGAAAGCAAGGTGCAACGCTGTGATAGAG 540
    Db 965 CTTAGGAAGCACTGGGATCTGTCTCAGAAAGCAAGGTGCAACGCTGTGATAGAG 1024
    Qy 541 ATAATACATATGAGAGAAAGTGGTCAATAATCTAGTGTGATGATTTCAAGAGGAC 600
    Db 1025 ACAACATTTAGAGAGAAAGTGGTCAATAATCTAGTGTGATGATTTCAAGAGGAC 1084
    Qy 601 ACCCACTAGGCGCTACAGGCTTTGCTCAGTGTGCAAACTGCTGGCAGCTGAGAGGG 660
    Db 1085 ACCCACTAGGCGCAACAGGCTTTGCTCAGTGTGCAAACTGCTGGCAGCTGAGAGGG 1144
    Qy 661 AAGCCGAAAAGAGCAAGTTCCTGTGTGCAAAAGTGTGCTGTGNCATTAATTANGCAT 720
  
```

```

Db      1145 AAGCCGAAA--GAGCAGGTTCCCGGGGCAAGGCGCTCTGACACAACTTAGGCT 1202
Qy      721 TGGAGAACTGTGTGTGTAACACTTACAGATGGGATTTTCCCGAAGCCGCACTTCC 780
Db      1203 CGAGAGAGCTGTGTGTGTCACCTCTACAGATGGC--TTTTCGCCAGCTGCCACTCC 1260
Qy      781 TTTTAGACTCATCAAAATTGAAGCCNGTTCACCAAGCTCTGCAAGTATATGTTTAA 840
Db      1261 TTCAGAA--CGACCGAGTTTACGCTGCCA--CCAGCTTCGACAGGGATGATTTCAA 1316
Qy      841 NNNAAATCTNGTTTAAAGNGAGATTGAGAGAAACNTNAAAGAGGAGGAGGAGAA 900
Db      1317 GGCMAAAGCTCGCTTTAA--GAGATTTGAGAGAAAGACTTGA---AGAGAGGGAGAA 1370
Qy      901 ATTGTGAAAGAAAATTCGNGGGAATTTTTCCTTCAGGAGGAAANAATGCGCTTG 960
Db      1371 GTTCGTGAAGAAAATCGGTGCG---ATTTTTCCTTCAAGTGAAGATGAGCCCTGGA 1425
Qy      961 GGGGTAAAG 970
Db      1426 GGCAGAAAG 1435

RESULT 8
LOCUS   BC018384      2626 bp      mRNA      linear      ROD 20-SEP-2002
DEFINITION Mus musculus, steroid carrier protein 2, liver, clone MGC:11534
ACCESSION BC018384
VERSION   BC018384.1 GI:117390902
KEYWORDS  MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS   Strausberg, R.
TITLE      Direct Submision
JOURNAL    Submitted (03-DEC-2001) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics Office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
           Email: cgabbs-r@mail.nih.gov
           Tissue Procurement: Jeffrey Green M.D.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
           DNA Sequencing by: Baylor College of Medicine Human Genome
           Sequencing Center
           Center code: BCM-HGSC
           Web site: http://www.hgsc.bcm.tmc.edu/cdna/
           Contact: amg@bcm.tmc.edu
           Gunaratne, P.H., Garcia, A.M., Lu, X., Hulik, S.W., Hale, S.M.,
           Yoon, V.S., Kowik, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
           Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
Series: IRAK Plate: 17 Row: n Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
Location/Qualifiers
1..2626
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/db_xref="MGC:11534 IMAGE:3710340"
/tissue_type="Mammary tumor, C3(1)-Tag model, Infiltrating
ductal Carcinoma, 5 month old virgin mouse."

```

```

CDS
/codon_start=1
/product="steroid carrier protein 2, liver"
/protein_id="AAH18384.1"
/db_xref="GI:17390903"
/db_xref="locusID:20280"
/translation="MPSVALKSPRLRVFVVGKTKPKRGGENSRDYPDMAKEAQ
KALBDQIPYSAVEQACVGYGDSGQRAIYHSLGLTGIPIINWNNGSGTALF
MAHOLIQGLANCVIALGFEKMERGSIQKFSRDTTPDKHLEVLIDKXGSAHPITP
OMPGYAGKHEMKGTKEHFAKIGKMHKSVNNTVPOPOEYSLSEVMKSPVEDF
LTILOCTPSGAAAAILISSEFVQYGLQSKAVIVQEMWTDLPSPREKSLIKV
GDKSKKARCTEKSGLTLPNDVILTEHDCSVLELTILELGLCPESGGTIVDRG
DNTYGVKAVINPSGGLISKHPLATGAQCELELQWLRGEGKQVPGARVALDHL
GLGAVVLTLYRMGFPPEASFRTHQVAAATSSAGDPSKALVKELEKLEEBEQ
FVKIIGIFAFVKDGPGEKATWVDVKNKGSVLPSPDKADCTITMADSDLLALM
TGKMPGSAFPGKIKINGMLAKLQNLQPEKATL"

BASE COUNT      765 a      521 c      642 g      698 t
ORIGIN
Query Match      58.0%; Score 576.6; DB 10; Length 2626;
Best local Similarity 79.1%; Pred. No. 1.4e-133;
Matches 767; Conservative 0; Mismatches 184; Indels 19; Gaps 7;

Qy      1 CTCACCCAGTTCCTCTGATGTTGGGTATGCTGGAAGAAACATATGAAAAATATG 60
Db      519 CACATCCCATTACTCTCAGATGTTGGGTATGCTGGGAAAGAACATATGAAAAATATG 578
Qy      61 GAACAAAATTGAACATTTGGAAAATTTGATGGAATAATCATTAACATTCAGTTATA 120
Db      579 GAACAAAATTGAACATTTGGAAAATTTGATGGAATAATCATTAACATTCAGTTATA 638
Qy      121 ACCCGTATTCCTCCAGTTCAGATGATACAGTTTATGATGATGATGATGATGATGATG 180
Db      639 AAGATATTCCTCCAGTTCAGATGATACAGTTTATGATGATGATGATGATGATGATGATG 698
Qy      181 TTTTGTATTTTGTATCTTACATATGTTGTTCCACTTCAGATGCTGACAGACAA 240
Db      699 TTTTGTATTTTGTATCTTACATATGTTGTTCCACTTCAGATGCTGACAGACAA 758
Qy      241 TTTTGTATTTTGTATCTTACATATGTTGTTCCACTTCAGATGCTGACAGACAA 300
Db      759 TTTTGTATTTTGTATCTTACATATGTTGTTCCACTTCAGATGCTGACAGACAA 818
Qy      301 TGGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Db      819 TGGCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 878
Qy      361 TGGTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
Db      879 TGGTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 938
Qy      421 CACCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
Db      939 CACCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 998
Qy      481 CTTATGAAGCACTGGGACTGTCTCCAGAGGCAAGAGTGAACGCTGGTTGATAGAGAG 540
Db      999 CTTAGGAAGCACTGGGACTGTCTCCAGAGGCAAGAGTGAACGCTGGTTGATAGAGAG 1058
Qy      541 ATATATCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db      1059 ACAACACTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
Qy      601 ACCGACTGAGGCTTACAGTCTGTCTCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db      1119 ACCGACTGAGGCTTACAGTCTGTCTCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1178
Qy      661 AAGCCGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db      1179 AAGCCGAAA--GAGCAGGTTCCCGGGGCAAGGCGCTCTGACACAACTTAGGCT 1236

```

QY 721 TGGAGAACTGCTGTTGTAACACTACAGATGGGGTTTCCCGGAGCGCAGTTCC 780  
DB 1237 CGGAGAGGTGTGGTTGTACCCCTACAGATGGG--TTTTCCTCCAGAGCTCCAGCTCC 1294  
QY 781 TTTTGAACCTCAAAATTGAAGCGNGTTCACCAAGCTCTGCAAGTATGTTTAA 840  
DB 1295 TTACAGA--CGACACAGTTTCAAGCTCTCCCA--CGAGCTCTGCAAGGAGATGATCAA 1350  
QY 841 NNNAAATCTGTTTAAAGGATTTGAAGAGAAACNTNAAAGAGGAGGAGGAGACA 900  
DB 1351 GGCAAACTCTCTCTTTAA--GGAGATTGAGAAAGAGCTTGA----AGAGGAGGAGGACA 1404  
QY 901 ATTTTGAAGAAATAATGCGGGAATTTTCCCTTCAAGGAGGAGGAGGAGGAGGAG 960  
DB 1405 GTTCTGTGAAGAAATCGGTGGC----ATTTTGTCTTCAAGATGAGATGAGCCCTGGA 1459  
QY 961 GGGGTAAAG 970  
DB 1460 GGCAAGAGAG 1469

RESULT 9  
RATSCPXA  
LOCUS RATSCPXA 2571 bp mRNA linear ROD 27-APR-1993  
DEFINITION Rat sterol carrier protein-x mRNA.  
ACCESSION M57453.1 GI:206876  
VERSION M57453.1 GI:206876  
KEYWORDS alternative splicing; sterol carrier protein.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Mammalia; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2571)  
AUTHORS Seedorf,U., Raabe,M. and Aasmann,G.  
TITLE Cloning, expression, and nucleotide sequence of rat liver sterol  
JOURNAL carrier protein 2 cDNAs  
MEDLINE J. Biol. Chem. 266 (1), 630-636 (1991)  
PUBMED 91093192  
REFERENCE 2 (bases 1 to 2571)  
AUTHORS Seedorf,U., Raabe,M. and Aasmann,G.  
TITLE Cloning, expression, and Nucleotide Sequence of Mouse Sterol  
JOURNAL carrier protein x cDNAs and Structure of a Related Pseudogene  
MEDLINE Unpublished (1992)  
PUBMED  
COMMENT Original source text: R.norvegicus (strain Wistar) adult liver,  
CDNA to mRNA.

FEATURES  
source location/Qualifiers  
1..2571  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
/ligase\_type="liver"  
/dev\_stage="adult"  
1..2571  
/gene="SCP-2"  
/gene="SCP-2"  
/product="2.7 kb mRNA"  
/evidence="experimental"  
1..1953  
/gene="SCP-2"  
/product="2.1 kb mRNA"  
/evidence="experimental"  
19..1662  
/gene="SCP-2"  
/codon\_start=1  
/product="sterol carrier protein-x"  
/protein\_id="AAA42122.1"  
/db\_xref="GI:206877"  
/translation="MPSVALNSPLRVFVVGVMTKPKPKGKNSRDYPLAKRAGQ

KALADROIPIYSAVEOACVGYVYGEESTCGORAIYHSLGTGPIIINNNCSNGSTALF  
MAOOLVOGLANCTIALGFEKMEKSGISCTKXSDNSNPLEKXIDIVLINYKMSACFPAP  
QIFSGAGKEHNETYGTAKYEHAKIGKMKRKHKSNNPNYSQFODESILDEIMSRPFPD  
LTVIQCCPTSDGAAAIYSSSEFQKHDSKAEIVAVQEMVTDMPTEEFESVTKMV  
GDYMSKEARACYERKSGSPDQVIVIELHDFSTNELTTEALGICLPGOGALVDRG  
DNTGSKAVINPSGLISKGPLAGTGAOCEALCQMRGEGAKQVGAQALQHLN  
GLGGAAYVTLRYMPEASSPFRTHQISAAFTSAGDSFKNLIPKEIEKLEEGEB  
FVKXIGIFARKVVDGPGKREATVYVYKNGKSVLPSPDKKADCTIMABSDLLAM  
TKRNPQSAFQGIKLIAGNGLAMKLSLOLPDKAKL"

BASE COUNT 724 a 524 c 642 g 681 t  
ORIGIN

Query Match 57.9%; Score 575.2; DB 10; Length 2571;  
Best Local Similarity 78.5%; Pred. No. 3.2e-133;  
Matches 758; Conservative 0; Mismatches 188; Indels 19; Gaps 6;

QY 6 CCAGTTGCTCTCAGATGTTGGTATGCTGGAAAAGAAATATGAAATATGAAACA 65  
DB 487 CCATTGCTCTCAGCTGTTGGAGCGCTGGAAAAGAAATATGAAATATGAAACA 546  
QY 66 AAATTTGAACACTTTGCAAAAATTGATGAGAAAATCTAAACATTCAATTATACCGG 125  
DB 547 AAGGTTGAACACTTTGCAAAAATTGATGAGAAAATCTAAACACTGTTATATACCGG 606  
QY 126 TATTCCTGCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185  
DB 607 TATTCCTGCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666  
QY 186 GATTTTGTGATCTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 245  
DB 667 GATTTTGTGATCTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 726  
QY 246 GCCAGTGAAGACTTTGTAACAGAGATGATGATGATGATGATGATGATGATGATGAT 305  
DB 727 TCTAGTGAAGACTTTGTAACAGAGATGATGATGATGATGATGATGATGATGATGAT 786  
QY 306 CAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 365  
DB 787 CAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 846  
QY 366 GGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 425  
DB 847 GGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 906  
QY 426 AATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 485  
DB 907 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 966  
QY 486 GAAAGACTGGGACTCTGTCCAGAAAGAGAAAGTGAACGCTGTTGATGAGAGATAT 545  
DB 967 GAAAGACTGGGACTCTGTCCAGAAAGAGAAAGTGAACGCTGTTGATGAGAGATAT 1026  
QY 546 ACATATGAGAGAAAGTGGGTATATAATCTGATGATGATGATGATGATGATGATGATGAT 605  
DB 1027 ACTTACGAGAGAAAGTGGGTATATAATCTGATGATGATGATGATGATGATGATGATGAT 1086  
QY 606 CTAGGCGCTACAGAGTCTGTGCTGAGTGCAGAACTCTGCTGAGCTGAGAGGAGAGCC 665  
DB 1087 CTGGGTGCTACAGAGTCTGTGCTGAGTGCAGAACTCTGCTGAGCTGAGAGGAGAGCC 1146  
QY 666 GGAAGAGAGGAGAAAGTCTGTGCTGAGTGCAGAACTCTGCTGAGCTGAGAGGAGAGCC 725  
DB 1147 GGAAGAGAGGAGAAAGTCTGTGCTGAGTGCAGAACTCTGCTGAGCTGAGAGGAGAGCC 1204  
QY 726 GAACTGTGTTGTAACACTTACAGATGAGGATTTTCCCGAAGCGCAGATTCCTTTTA 785  
DB 1205 GAGCTGTGTTGTAACACTTACAGATGAGGATTTTCCCGAAGCGCAGATTCCTTTTA 1261  
QY 786 GAACTGTGTTGTAACACTTACAGATGAGGATTTTCCCGAAGCGCAGATTCCTTTTA 845  
DB 1262 GAACTGTGTTGTAACACTTACAGATGAGGATTTTCCCGAAGCGCAGATTCCTTTTA 1318  
QY 846 AATCTNGTTTAAAGGAGATTTGAGAGAGAAACNTNAAAGGAGGAGGAGGAGAACTTTG 905



Db 1319 ATCTCATTTTAAAGAA-----ATCGAAGAAAGCTTGAAGAGAGAGGAAAGCTTG 1372  
Qy 906 TGAAGAAAAATNCGGAGAAATTTTTCCTTCAAGGGGAAATGCGCTCGGGGGT 965  
Db 1373 TGAAGAAAAATCGGTGC-----ATTTTTCCTTCAAGAGAGATGAGCGCGGGGCA 1427  
Qy 966 AAAAG 970  
Db 1428 AGAAG 1432

RESULT 10  
RAT60KDA  
LOCUS  
DEFINITION Rat 60 kDa protein and non-specific lipid transfer protein mRNA,  
complete cds.  
ACCESSION M62763  
VERSION 60 kDa protein, non-specific lipid transfer protein.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2599)  
AUTHORS Mori,T., Tsukamoto,T., Mori,H., Tashiro,Y. and Fujiki,Y.  
TITLE Molecular cloning and deduced amino acid sequence of non-specific  
lipid transfer protein (sterol carrier protein 2) of rat liver: a  
higher molecular mass (60 kDa) protein contains the primary  
sequence of non-specific lipid transfer protein as its C-terminal  
part  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (10), 4338-4342 (1991)  
MEDLINE 91239563  
PUBMED 2034675  
COMMENT Original source text: Rattus norvegicus (strain Sprague-Dawley)  
cDNA to mRNA.  
FEATURES  
source location/Qualifiers  
1..2599  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
12..1977  
/product="clone TM-34"  
46..1689  
/standard\_name="clone TM-34"  
/codon\_start=1  
/product="60 kDa protein"  
/protein\_id="AAA40622.1"  
/db\_xref="GI:202553"  
/translation="MPSVALNSPRLRFVVGVMTKFKPGGNSRDYDILAKAQQ  
KALDAQIPIYSAVEQACVGVYGEISCTGQRAIYHSLGLGIRIINNNSCTGSLAF  
MAQOUVPGSLANCVLALGEEKMEKSLGTRYKSDRSNPLEKHLIDVINKMGACFPAP  
QIFGSGKHEMTYGTGVYHFAKIKGNHSHVNNPYSOFOEDYSIDETMKRPYED  
LTVLCCPTSDGAATAIVSSEFVKGKOSKAVETVQEMTMDPSTPEESVTKM  
GYDMSEAKRKCYEKSGLSPDYVIELHDCSTNELTYELTGICPEBGALVDRC  
DNTYGGKWINSPGGLISKGHPUGATGALQCAELCWQLGEGKQVPAKVALDHL  
GGGAVALVLYRMGPPEASFRTHQISAPPSAGDGPKANLIFREIKLEBEGEE  
FYKIKGIFAFKVDGPGGKEATWVDVNGKGSVLPDSKADCTTTWADSLALM  
TGKMPDFAFPGKLIKAGNMLAKMLQSLQDPKAKL"  
1258..1689  
/note="larger precursor of non-specific lipid transfer  
carrier protein (sterol carrier protein 2)"  
/codon\_start=1  
/product="non-specific lipid transfer protein"  
/protein\_id="AAA40623.1"  
/db\_xref="GI:202554"  
/translation="MGFPAASSFRTHQISAPPSAGDGPKANLIFREIKLEBEG  
EEFYKIKGIFAFKVDGPGGKEATWVDVNGKGSVLPDSKADCTTTWADSLALA  
LMTGKMPDFAFPGKLIKAGNMLAKMLQSLQDPKAKL"  
1258..1317  
1318..1686  
sig\_peptide  
mat\_peptide

/product="non-specific lipid transfer protein"  
/note="mature non-specific lipid transfer carrier protein  
(sterol carrier protein 2)"  
polyA\_signal  
1953..1958  
/note="clone TM-11 & 34"  
polyA\_signal  
2160..2165  
/note="clone TM-28 & 33"  
polyA\_signal  
2242..2247  
/note="clone TM-28 & 33"  
polyA\_signal  
2382..2387  
/note="clone TM-28 & 33"  
polyA\_signal  
2580..2585  
/note="clone TM-33"  
polyA\_signal  
2581..2586  
/note="clone TM-33"  
BASE COUNT 727 a 524 c 654 g 694 t  
ORIGIN  
Query Match 57.9%; Score 575.2; DB 10; Length 2599;  
Best Local Similarity 78.5%; Pred. No. 3.2e-133;  
Matches 758; Conservative 0; Mismatches 188; Indels 19; Gaps 6;  
Qy 6 CCAGTTGCTCTCAGATGTTGGTATGCTGAGAAAAAATATGAGAAACATATGAGAA 65  
Db 514 CCATTGCTCTCAGATGTTGGTATGCTGAGAAAAAATATGAGAAACATATGAGAA 573  
Qy 66 AAAATGAACATTTTGCAAAAATTTGATGAGAAAAATCTAAACATTCATTAAACCG 125  
Db 574 AAGGTGAACATTTTGCAAAAATTTGATGAGAAAAATCTAAACATTCATTAAACCG 633  
Qy 126 TATTCCTCAAGTTCAGATGATACAGTTAGATGAGATGATGATGATGATGATGATG 185  
Db 634 TATTCCTCAAGTTCAGATGATACAGTTAGATGAGATGATGATGATGATGATGATG 693  
Qy 186 GATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 245  
Db 694 GATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 753  
Qy 246 GCCAGTGAAGCATTTGTACAGAAAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 305  
Db 754 TCTAGTGAAGCATTTGTACAGAAAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 813  
Qy 306 CAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 365  
Db 814 CAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 873  
Qy 366 GCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 425  
Db 874 GCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 933  
Qy 426 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 485  
Db 934 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 993  
Qy 486 GAAAGACTGAGACTCTGTCAGAAAGACAAGGTGCAACGCTGATGATGAGAGATATAT 545  
Db 994 GAAAGACTGAGACTCTGTCAGAAAGACAAGGTGCAACGCTGATGATGAGAGAGATATAT 1053  
Qy 546 ACATATGAGAGAAAGTGGGCTATTAATCTCTAGTGTGAGCTGATTTCAAGGAGACACCA 605  
Db 1054 ACTTACGAGAGAAAGTGGGCTATTAATCTCTAGTGTGAGCTGATTTCAAGGAGACACCA 1113  
Qy 606 CTAGGCGTACAGGCTTGTGCTTCAAGTGTGCAAGAACTCTGCTGGCACTGAGAGGAGAGCC 665  
Db 1114 CTAGGCGTACAGGCTTGTGCTTCAAGTGTGCAAGAACTCTGCTGGCACTGAGAGGAGAGCC 1173  
Qy 666 GGAAGAAAGAGCAAGGCTTGTGCTTCAAGTGTGCAAGAACTCTGCTGGCACTGAGAGGAGAGCC 725  
Db 1174 GGAAGAAAGAGCAAGGCTTGTGCTTCAAGTGTGCAAGAACTCTGCTGGCACTGAGAGGAGAGCC 1231  
Qy 726 GAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 785  
Db 1232 GAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1288



```

QY 786 GAACATCATGAATTGAAGCCNGTTCACACCAAGCTCTGCAAGTATGATGTTTAAANGNA 845
DB 1289 GAACGACACCAATTTCA---GCTGCTCCACAGCTCTCAGGGAGTGAATTCAGGCA 1345
QY 846 AATCTNGTTTAAAGNGGATTTAGAGAGAACTNNAAGAGGAGGAGCAATTTG 905
DB 1346 ATCTCATTTTAAAGAA-----ATCGAAGAGAGCTTGAAGAGAGGAGGAGGTTG 1399
QY 906 TGAAGAAATNCGNGGAAATTTTGCCTTCAAGGGGAAANAATGCGCTGGGGGT 965
DB 1400 TGAAGAAATCGGTGC-----ATTTCCTTCAAGTGAAGAGATGCGCGGGGGCA 1454
QY 966 AAAAG 970
DB 1455 AGAAG 1459

RESULT 11
AX306091 2152 bp DNA 1linear PAT 11-DEC-2001
LOCUS AX306091
DEFINITION Sequence 842 from Patent WO01818188.
ACCESSION AX306091
VERSION AX306091.1 GI:17645404
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Ishikawa,K., Arai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
AUTHORS Method for examining ischemic conditions
TITLE Patent: WO 0181818-A 842 22-NOV-2001;
JOURNAL School Juridical Person Nihon University (JP)
FEATURES
source
1..2152
location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

BASE COUNT 586 a 455 c 576 g 535 t
ORIGIN

Query Match 57.7%; Score 573.4; DB 6; Length 2152;
Beet Local Similarity 78.9%; Pred. No. 9.1e-133;
Matches 765; Conservative 0; Mismatches 186; Indels 19; Gaps 7;

QY 1 CTCACCCAGTTGCTCTCAGATGTTTGGGTATGCTGGAAGAACAATGAAAAATATG 60
DB 479 CACATCCGATTACTCTCAGATGTTTGGGTATGCTGGAAGAACAATGAAAAATATG 538
QY 61 GAACAAAAATTTGACACTTTGCAAAAAATTTGATGAAAAATCATTAACATTCAATTATA 120
DB 539 GAACAAAAATTTGACACTTTGCAAAAAATTTGATGAAAAATCATTAACACTCAATTATA 598
QY 121 ACCCGATTTCCAGTTCCAAAGATGAATACAGTTTAAAGAGATGAGATCAAAAGAG 180
DB 599 AACAGTATTCAGATTCCAAGATGAATACAGTTTAAAGAGATGAATGAATCAAAACGAG 658
QY 181 TTTTGAATTTTGAATCTTACATGTTTGTCCCACTTCAGATGCTGTCAGCAGCA 240
DB 659 TTTTGAATTTTGAATCTTACATGTTTGTCCCACTTCAGATGCTGTCAGCAGCA 718
QY 241 TTTTGGCCAGTGAAGCATTTGTACAGAGATGAGCTGTCATTCAAAGCTGTGGAATTT 300
DB 719 TTTCTGTCAGCAGAGGAGTTGTGACAGACAGCTGACAGTCCAAAGCGGTGAGATTG 778
QY 301 TGGCACAAGAAATGATGATGATTTGCAAGCTCGTTTGAAGAAAAAGCATTTATTA 360
DB 779 TGGCCCAAGAGATGATGATGATTTGCAAGCTCGTTTGAAGAAAAAGCATTTATTA 838
QY 361 TGGTGGCTTTGATGATGATTAAGAGAGCTGCAAGAAATGCTATGAGAAATCTGAGCTGA 420
DB 839 TGGTGGCTATGATGATGATTAAGAGAGCTGCAAGAGATGCTATGAGAAATCTGAGCTGA 898

```

```

QY 421 CACCAATGATATTGACGTAATAGAACTTCAAGATTTCTTTCTACCAAGAACTCCTTA 480
DB 899 CACCAAGAGATGTCAGCGTAGAGACTTCAAGATTTCTTTCTCTCAATGAACTCATCA 958
QY 481 CTATGAAGCACTGGAGCTCTGTCCAGAGAGCAAGTGAACAGCTGTTGATAGAGAG 540
DB 959 CTATGAAGCACTGGAGCTCTGTCCAGAGAGCAAGTGAACAGCTGTTGATAGAGAG 1018
QY 541 ATATATCATATGAGAGAAAGTGGGTCTATTAATCTTATGTTGAGTCAATTTCAAGGAC 600
DB 1019 ACAACACTTACGAGAGAAAGTGGGTCTATTAATCTTATGTTGAGTCAATTTCAAGGAC 1078
QY 601 ACCCATGAGGCTTCAAGTCTGTCTCAGTGTGAGAGAACTCTGTCGAGCTGAGAGAGG 660
DB 1079 ACCCATGAGGCTTCAAGTCTGTCTCAGTGTGAGAGAACTCTGTCGAGCTGAGAGAGG 1138
QY 661 AAGCCGAGAAAGAGCAAGTCTCTGTGCAAGAGTGGCTCTGTCGATTAATTTANGCAT 720
DB 1139 AAGCCGAGAA--GAGGAGGTTCCCGGGGCAAGGTGGCTCTGTCGATTAATTTANGCAT 1196
QY 721 TGGAGAACTGTGTTGTTGACACTTACAGATGAGGTTTTCGCGAAGCCGCAAGTTC 780
DB 1197 CGAGAGAGCTGTGTTGTTGACACTTACAGATGAGG--TTTTCGCGAAGCTCCAGCTCC 1254
QY 781 TTTTGAATCATCAAAATTTGAAGCCNGTTCACACCAAGCTCTGCAAGTATNGTTTAA 840
DB 1255 TTCAAGA--CGCACAGGTTTCAAGTCTGCTCCCA--CAAGCTCTGCAAGGAGATGATTCAN 1310
QY 841 NGNAATATCTNGTTTAAAGNGATTTGAAGAGAACTNNAAGAGGAGAAAGGAGACA 900
DB 1311 GGCAAACTCGCTTTA--GAGATTTGAAGAGAAAGCTTGA--AGAGAGAGGAGACA 1364
QY 901 ATTGTGAAAGAAATNCGNGGAAATTTTTCCTTCAAGGGGAAANAATGCGCTTGG 960
DB 1365 GTTCGTGAAGAAATCGGTGC-----ATTTCCTTCAAGTGAAGAGATGAGCTCGGA 1419
QY 961 GGGGTAAAG 970
DB 1420 GGCAAGAGAG 1429

RESULT 12
MUSSCP 2152 bp mRNA 1linear ROD 14-JUL-1993
LOCUS MUSSCP
DEFINITION Mus musculus sterol-carrier protein X mRNA, complete cds.
ACCESSION M91458
VERSION M91458.1 GI:293793
KEYWORDS non-specific lipid transfer protein; sterol carrier protein.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Seedorf,U., Raabe,M. and Aegmann,G.
AUTHORS Cloning, expression and sequences of mouse sterol-carrier
TITLE protein-x-encoding cDNAs and a related pseudogene
JOURNAL Gene 123 (2), 165-172 (1993)
MEDLINE 93154580
PUBMED 8428655
COMMENT Original source text: Mus musculus (strain C57BL/ 6 x CBA)
(library: 1-ZAP II) female adult, 6-8 weeks liver cDNA to mRNA.
location/Qualifiers
1..2152
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/ 6 x CBA"
/db_xref="taxon:10090"
/bx="female"
/tissue_type="liver"
/dev_stage="adult, 6-8 weeks"
/tissue_lib="1-ZAP II"
16..1659
CDS

```

```
/codon_start=1
/product="sterol-carrier protein X"
/protein_id="AA40098.1"
/db_xref="GI:293794"
/translation="MPSVALYKSPRLRRVFGVGMKPKPGGNSRDYPDMAKENGQ
KLELDQIPIYSAVEACVGVYGDSTSGRAIYHSLGLGIPLINNNCISGSLTF
MAHOLIQGLANCVLAFGEKMERGSIGKFSRPTPTDKHLEVIDKGLAMPTLP
OMPGYAGKEHMEKYGKVPKIKGKHSHSVNNTYSSQPODEYSLSEYMKSRVPDF
LTILOCCPTSDACAIILSSEFVOYGIQSVAVELVADEMPTDSTTEESKTIKV
GYDMSKEARRCYKESGLTPNDVDVLEHDCFSVNLITYELGLCPBGQSTLVDRG
DNYGKAIINPSGGLISKQPLGATGACALCWQLHGEAGKRVPGAKVALQNL
GIGAVVTLRYMGPPEAASFRTHOVSAAPTSASADGKAMLVFEIEIKLEEBEQ
FKKIGIIPAFKVDGPGKEATWVDVNGKGSYVPSNDKXADCTTTMADBDLLALM
NGKMPQSAFPGQKLIKINMGLAMQLQLOPKAKL"
polya_signal
polya_site 2137..2142
BASE COUNT 586 a 455 c 576 g 535 t
ORIGIN
Query Match 57.7%; Score 573.4; DB 10; Length 2152;
Best Local Similarity 78.9%; Pred. No. 9.1e-133;
Matches 765; Conservative 0; Mismatches 186; Indels 19; Gaps 7;
QY 1 CTCACCCAGTTGCTCCTCAGATGTTGGTATGCTGAAAAAGAACATATGAAAAATATG 60
DB 479 CACATCCGATTACTCTCAGATGTTGGGTATGCTGGAAGAACAATATGAAAAATATG 538
QY 61 GAACAAAAATTGAACACTTTGCAAAAAATTGATGAAAAATCATAAACATTCAGTTAATA 120
DB 539 GAACAAAAATTGAACACTTTGCAAAAAATTGATGAAAAATCATAAACATTCAGTTAATA 598
QY 121 ACCCGATTCACAGTCCAAAGATGATGATTAAGTGAAGGATGATCAAAAGAG 180
DB 599 AACAGTATTCACAGTCCAAAGATGATGATTAAGTGAAGGATGATCAAAAGAG 658
QY 181 TTTTGTATTTTGAATCTTACATGTTGTCCTCACTTCAGATGCTGTCAGCAGCA 240
DB 659 TTTTGTATTTTGAATCTTACATGTTGTCCTCACTTCAGATGCTGTCAGCAGCA 718
QY 241 TTTTGGCCAGTGAACATTTGTAACAGATGAGCCTGCAATCCAAAGCTGTGAAATTT 300
DB 719 TTTGTGCGACGAGGAGTTGTGACAGATGAGCCTGCAATCCAAAGCTGTGAAATTT 778
QY 301 TGGCAACAAGAAATGATGATGATTTGCAAGCTCGTTGAAGAAAAAGCATTTATTA 360
DB 779 TGGCCAGAGATGATGATGATTTGCAAGCTCGTTGAAGAAAAAGCATTTATTA 838
QY 361 TGGTGGCTTTGATATGATGATTAAGAGAGTGAAGAAATGCTATGAGAAATCTGCGCTGA 420
DB 839 TGGTGGCTATGATATGATGATTAAGAGAGTGAAGAAATGCTATGAGAAATCTGCGCTGA 898
QY 421 CACCAATATATGATGATGATTAAGAGAGTGAAGAAATGCTATGAGAAATCTGCTTA 480
DB 899 CACCAATATGATGATGATTAAGAGAGTGAAGAAATGCTATGAGAAATCTGCTTA 958
QY 481 CTATGAACACTGGGAGCTGTGCGAGAGAGCAAGGTCAACGCTGTGATGAGAGAG 540
DB 959 CTATGAACACTGGGAGCTGTGCGAGAGAGCAAGGTCAACGCTGTGATGAGAGAG 1018
QY 541 ATAAATCATATGAGAGAGAGTGTGATTAATCTATGATGATGATGATTTGAAAGGAG 600
DB 1019 ACAACACTTACGAGAGAGAGTGTGATTAATCTATGATGATGATGATTTGAAAGGAG 1078
QY 601 ACCCACTAGCGGTACAGGTCTTGCTCAGTGTGACAGACTTGCTGTGAGAGAGAG 660
DB 1079 ACCCACTAGCGGTACAGGTCTTGCTCAGTGTGACAGACTTGCTGTGAGAGAGAG 1138
QY 661 AAGCCGGAAGAGAGAGAGTCTGTGTCGAAAGAGTCTGTGTCGTCGTCGTCGTCGTCG 720
DB 1139 AAGCCGGAAGAGAGAGAGTCTGTGTCGAAAGAGTCTGTGTCGTCGTCGTCGTCGTCG 1196
QY 721 TGGAGAACTGTGTTGTAACACTTACAGAGATGGGGTTTCCCGGAACCGCGAGTTC 780
```

```
DB 1197 CGAGAGAGCTGTGTTGTCACCCCTCTACAGATGG--TTTCCCGAAGCTGCAGCTCC 1254
QY 781 TTTTGAACATCTCAAAATTGAAGCCNGTTCACCAAGCTTGCAAGTATNGSTTAA 840
DB 1255 TTCAAGAA--CGCACAGAGTTTCAGCTGCTCCCA--CCAGCTCTGAGAGGAGATTCA 1310
QY 841 NNNAAATCTGTTTAAAGGAGATGGAAGAAACNTAAAGAGGAAGGGAACA 900
DB 1311 GGCACCACTCGCTTTAA--GAGAGTTGAGAGAAAGCTTGA---AGAGGAAGGGAACA 1364
QY 901 ATTGTGAAAGAAATATGCGNGGGAATTTTTCCTTCAAGGGGAAANAATGAGCCTGG 960
DB 1365 GTTCGTGAAGAAATATCGGTGGC-----ATTTTGTCTTCAAGTGAAGATGAGCCTGGA 1419
QY 961 GGGGTAAAG 970
DB 1420 GGCAGAAAG 1429
RESULT 13
RATSCP2
LOCUS RATSCP2 1409 bp mRNA linear ROD 27-APR-1993
DEFINITION Rat sterol carrier protein-2 (SCP-2) mRNA, complete cds.
ACCESSION M34728
VERSION M34728.1 GI:206871
KEYWORDS sterol carrier protein-2.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1409)
AUTHORS Billheimer,J.T., Strehl,L.L., Davis,G.L., Strauss,J.F. III and
Davis,L.G.
TITLE Characterization of a cDNA encoding rat sterol carrier protein2
JOURNAL DNA Cell Biol. 9 (3), 159-165 (1990)
MEDLINE 90253610
PUBMED 2340090
COMMENT Original source text: Rat liver, cDNA to mRNA, clone SP43.
FEATURES
source
1..1409
location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
<1..1409
/product="sterol carrier protein-2 mRNA"
307..1128
/notes="sterol carrier protein-2 (SCP-2)"
/codon_start=1
/protein_id="AA42120.1"
/db_xref="GI:206872"
/translation="WVGIDMSKEARKCYEAGSPVDVLEHDCSTBELTYEA
LGLCPGEGGALVIRGDNVTYGGKWINSGGLISKHPLGATGLAQCELQWLRGEA
GKQVPGAKVALQNLGIGAAVVTLYMGPPEAASFRTHQISAPTSASADGKAM
LVFEIEIKLEEBEYFKIGIIFAFKVDGPGKEATWVDVNGKGSYVPSNDKX
ADCTTMAADBDLLALMNGKMPQSAFPGQKLIKINMGLAMQLQLOPKAKL"
BASE COUNT 387 a 300 c 390 g 332 t
ORIGIN
Query Match 52.5%; Score 522.2; DB 10; Length 1409;
Best Local Similarity 77.8%; Pred. No. 6.1e-120;
Matches 713; Conservative 0; Mismatches 183; Indels 20; Gaps 7;
QY 55 AATATGAACAAAAATTGAACACTTTGCAAAAAATTGATGGAANAATATATAAATTGAG 114
DB 3 AATTCGGAACAAAGGTTGAACACTTTGCAAAAAATTGATGGAANAATATATAAACACTG 62
QY 115 TTAATACCCGATTCACAGTTCAGATGAATACGTTTGAATGAAGTGAAGTGCATCTA 174
DB 63 TTAATACCCGATTCACAGTTCAGATGAATACGTTTGAATGAAGTGAAGTGAATGAATCA 122
QY 175 AAGAAGTTTGAATTTTGAATCTTACATGTTGTCACATGTTGCCACTTCAGATGATGCTGAG 234
```

Db 123 GGCACATTTTCGATTTTCTGACTGCTCTTACAAATGCTGTCCCACTCAGATGGTCCGACG 182  
 Qy 225 CAGCAATTTTGGCCAGTGAAGCATTTTGTACAGAAATGAGCTTGCATTCGAAAGCTGTGG 294  
 Db 183 CAGCAATTTTGTCTAGTGTGAGGATTTGTGCAAGATGAGCTGTGACAGTCCAAAGCTGTGG 242  
 Qy 295 AAATTTTGGCACAAGAAATGATGACTGATTTTGCAGATTCGTTTGAAGAAAGACATTA 354  
 Db 243 AAATTTTGGCACAAGAAATGATGACTGATTTTGCAGATTCGTTTGAAGAAAGACATTA 301  
 Qy 355 TTAATAATGTTGGCTTTGATATGATGATTAAGAAAGCTGCAAGAAATGCTATGAGAAATCTG 414  
 Db 302 TTAATAATGTTGGCTTTGATATGATGATTAAGAAAGCTGCAAGAAATGCTATGAGAAATCTG 361  
 Qy 415 GCTTACACCAATATGATATGATGATTAAGAAAGCTGCAAGAAATGCTATGAGAAATCTG 474  
 Db 362 GCTTACACCAATATGATATGATGATTAAGAAAGCTGCAAGAAATGCTATGAGAAATCTG 421  
 Qy 475 TCCCTACTATTAAGACACTGGGACTGTCTCCAGAAAGCAAGTGCACGCTGTGATTA 534  
 Db 422 TCCCTACTATTAAGACACTGGGACTGTCTCCAGAAAGCAAGTGCACGCTGTGATTA 481  
 Qy 535 GAGGAGATTAATACATATGAGAGAAAGTGGGTATTAATCTAGTGTGACTGATTTCA 594  
 Db 482 GAGGAGATTAATACATATGAGAGAAAGTGGGTATTAATCTAGTGTGACTGATTTCA 541  
 Qy 555 AGGAGACCCCACTAGGCGCTTACAGGCTTGTCTCAGTGTGACAACTCTCTGCGACGCTGA 654  
 Db 542 AGGAGACCCCACTAGGCGCTTACAGGCTTGTCTCAGTGTGACAACTCTCTGCGACGCTGA 601  
 Qy 655 GAGGGAGAGCGGAAAGAGGCAAGTCTGTGTGCAAGAGTGGCTCTGCGACGCTGA 714  
 Db 602 GAGGGAGAGCGGAAAGAGGCAAGTCTGTGTGCAAGAGTGGCTCTGCGACGCTGA 659  
 Qy 715 AAGCATTTGAGAGAACTGTGTGTATACACTTACAGATGAGGATTTTCCGGAAGCGCC 774  
 Db 660 AAGCATTTGAGAGAACTGTGTGTATACACTTACAGATGAGGATTTTCCGGAAGCGCC 717  
 Qy 775 AGTCTCTTTAAGACTCATCAAAATTTGAAGCCGTTCCAAACCAAGCTCTGCAAGTATATG 834  
 Db 718 AGCTCC-TTCAGAACGACCAAGATTTCA---GCTGCTCCACACGCTCTGCGAGGAGATG 773  
 Qy 835 GTTTAANGAATCTNGTTTAAAGGNGATTGAGAGAAACNTNMAAGGGAAGG 894  
 Db 774 ATTCAGAGCAATTCATTTTAAAGAA-----ATCGAAGAAAGCTTGAAGAGAGAG 827  
 Qy 895 GGAACATTTTGTGAAGAAATTCGNGGGAATTTTGCCTTCAAGGGAAGAAATATG 954  
 Db 828 GGAAGAGTTTGTGAAGAAATTCGNGGGAATTTTGCCTTCAAGGGAAGAAATATG 882  
 Qy 955 CCCTGGGGGGTAAAG 970  
 Db 883 CCCGGGGGCAAGAAAG 898  
 RESULT 14  
 H0MSCP2B  
 LOCUS H0MSCP2B 1500 bp mRNA linear PRI 06-DEC-1993  
 DEFINITION Humn sterol carrier protein 2 mRNA, complete cds.  
 ACCESSION M75884  
 VERSION M75884.1 GI:432976  
 KEYWORDS sterol carrier protein-2.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 AUTHORS He,Z., Yamamoto,R., Furch,E.E., Schantz,L.J., Naylor,S.L.,  
 George,H., Billheimer,J.T. and Straube,J.F., III.  
 TITLE cDNA encoding members of a family of proteins related to human  
 sterol carrier protein 2 and assignment of the gene to human  
 chromosome 1 p21----pter  
 JOURNAL DNA Cell Biol. 10 (8), 559-569 (1991)

MEDLINE 92029618  
 PUBMED 1718316  
 REFERENCE 2 (bases 1 to 1500)  
 AUTHORS Vesa,J., Hellesten,E., Branowski,B.L., Emanuel,B.S., Billheimer,J.T.,  
 Mead,S., Cowell,J.K., Straube,J.F.,III, and Pellonem,L.  
 TITLE Assignment of sterol carrier protein X/sterol carrier protein 2 to  
 1p32 and exclusion as the causative gene for infantile neuronal  
 ceroid lipofuscinosis  
 JOURNAL Unpublished  
 COMMENT On Dec 6, 1993 this sequence version replaced gi:337996.  
 FEATURES  
 source Human liver cDNA to mRNA.  
 location/Qualifiers  
 1..1500  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /map="1p32"  
 /tissue\_type="liver"  
 /clone\_id="Clontech HL-1001b/HL-1115b"  
 1..1500  
 /gene="SCP-2"  
 82..951  
 /gene="SCP-2"  
 /codon\_start=1  
 /product="sterol carrier protein-2"  
 /protein\_id="AA03558.1"  
 /db\_xref="GI:432977"  
 /translation="MTDLPSFEEKSIIRKVGFDMSKEARKCYEKGSLPNDIDVI  
 ELHDFSTNELLTTEALGLCEGGGATLVLDGDNVYGGKWIINPSGLISKHPIGAT  
 GLAOCALBQMLRGAGRGVPAKALVQHNHIGAGAVVTLIKMGFPELAASSPFTQ  
 IEAVPTSSASDGFPAKILVFEIKLEKEEGOPFYKIFAFKVDKDPGGEATWV  
 DVKNGKGSVLPNSDKKADCTITMADSDPLALMTGMNPOSAFPOGKLITNGMLAMK  
 LQNLQDQPKAKL"  
 1218..1223  
 /gene="SCP-2"  
 /note="no polyA or polyA signal was found at the end of  
 this cDNA clone, indicating that the mRNA sequence may be  
 partial at the 3' end"  
 /evidence="not experimental"  
 BASE COUNT 482 a 254 c 342 g 422 t  
 ORIGIN  
 Query Match 51.1%; Score 508.2; DB 9; Length 1500;  
 Best Local Similarity 87.8%; Pred. No. 2e-116;  
 Matches 640; Conservative 0; Mismatches 69; Indels 20; Gaps 8;  
 Qy 257 ATTGTACAGAAATGAGCTGTGCAATCCAAAGCTGTGAAATTTTGGCACAAGAAATGAT 316  
 Db 27 ATTTTACAGAAATGAGCTGTGCAATCCAAAGCTGTGAAATTTTGGCACAAGAAATGAT 86  
 Qy 317 GACTGATTTGCCAGCTGTTTGAAGAAAGAAAGCAATTAATAATGTTGGCTTTGATAT 376  
 Db 87 GACTGATTTGCCAGCTGTTTGAAGAAAGAAAGCAATTAATAATGTTGGCTTTGATAT 146  
 Qy 377 GAGTAAGAAGCTGCAAGAAATGCTATGAGAAATCTGGCCGACACCAAAATGATTTGA 436  
 Db 147 GAGTAAGAAGCTGCAAGAAATGCTATGAGAAATCTGGCCGACACCAAAATGATTTGA 206  
 Qy 437 CGTAATAGAACTTCAGATTTGTTTTCACCAAGCAATCTTACTTATGAAGACTGG 496  
 Db 207 CGTAATAGAACTTCAGATTTGTTTTCACCAAGCAATCTTACTTATGAAGACTGG 266  
 Qy 497 ACTTGTCCAGAGCAAGGTCACAGCTGTGTTATAGAGATTAATACATATGAGAG 556  
 Db 267 ACTTGTCCAGAGCAAGGTCACAGCTGTGTTATAGAGATTAATACATATGAGAG 326  
 Qy 557 AAAGTGGGTCAATTAATCTAGTGTGAGCTGATTTCAAGGAGCAACCACTAGGGGCTAC 616  
 Db 327 AAAGTGGGTCAATTAATCTAGTGTGAGCTGATTTCAAGGAGCAACCACTAGGGGCTAC 386  
 Qy 617 AGGCTTCTCAGTGTGAGAACTGTGTGCGAGCTGAGAGGAGGACCGGAAAGAGAGC 676  
 Db 387 AGGCTTCTCAGTGTGAGAACTGTGTGCGAGCTGAGAGGAGGAGGAAAGCGGAA--GAGG 444

```

QY 677 AAAGTCTGTCGAAGGAGGCTGTCGNCATATTTANGCATTTGAGAGACTGCGTT 736
DB 445 CAAAGTCTGTCGAAGGAGGCTGTCGNCATATTTANGCATTTGAGAGACTGCGTT 504
QY 737 GTAACACTCTACAGATGAGGCTTTCCGGAAGCCGCACTTCTTTAGAACTCATCA 796
DB 505 GTTACACTCTACAGATGAGGCTTTCCGGAAGCCGCACTTCTTTAGAACTCATCA 561
QY 797 AATTGAGCGGCTTCCACCAAGCTCTGCAAGTATNGTTTANGNAAAATCTGCTTTT 856
DB 562 ATTGA--GCTGCTTCCACAGCTCTGCAAGTATNGTTTANGNAAAATCTGCTTTT 618
QY 857 AAAGGCGATGAGAGAAACNTNAAAGGGAAGGGAACAATTTGTGAAGAAAA 916
DB 619 AA--GGAGATGAGAGAACTTGA--AGAGAAAGGGAACAATTTGTGAAGAAAA 672
QY 917 TACGAGGGAATTTTCCCTTCAAGGGAANAATGCGCTGGGGGTAAAGANGGCC 976
DB 673 GGTGGT-----ATTTTGTCTTCAAGTGAAG-ATGCGCTGGGGGTAAAGANGGCC 726
QY 977 ACCCTGGGG 985
DB 727 TGGGTGGT 735

RESULT 15
CHSCP2A
LOCUS CHSCP2A 2443 bp mRNA linear VRT 11-AUG-1993
DEFINITION Gallus gallus domesticus sterol carrier protein-2 mRNA, 3' end.
ACCESSION L09231
VERSION L09231.1 GI:304422
KEYWORDS sterol carrier protein-2.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2443)
AUTHORS Pfeiffer,S.M., Sakiragci,N., Ryan,A., Johnson,A.L., Deeley,R.G.,
TITLE Billheimer,J.T., Baker,M.E. and Straus,J.F. III.
Chicken sterol carrier protein 2/sterol carrier protein x: cDNA
cloning reveals evolutionary conservation of structure and
regulated expression
JOURNAL Arch. Biochem. Biophys. 304 (1), 287-293 (1993)
MEDLINE 83312016
PUBMED 8323294
COMMENT Original source text: Gallus gallus (strain White Leghorn,
sub species domesticus) (library: lambda gt11) adult liver cDNA to
mRNA.

FEATURES
source
1.2443 Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn"
/sub_species="domesticus"
/db_xref="taxon:9031"
/issue_type="liver"
/dev_stage="adult"
/cisue_lib="lambda gt11"
<1..1646
/standard_name="SCP2"
/codon_start=3
/product="sterol carrier protein-2"
/protein_id="AA0248.1"
/db_xref="GI:304423"
/translation="ARVCEBGGPAAIMORVYVVGWMTKFAKSENSVDYDLAKE
AGOKALADGIPYSAVEACVGVVGDSTGCRATYHGLGLGPIINNNKCATGST
ALFMRQLVEGLADCVLALGFBRMAKSLAGFSDRTPMDKHLKIMNKYGLASAP
LTPQWVLANGKHEKERYGTNPEYFAKIAWKHSHSTNNPYSQFOKKYTLDEVLSRKV
FPLTLVLOCCPTNSMAAAILASDEPVRKHLQPOVAIVLQVMAITDPSTRENSCM
KMGYDMTKAAEKCKFKALPTDIVDELHDFSNVEFIYEAIGLCPEKACDLI
DRGDMYGGKWTINPSGGLISGHPILGATGLAQSALCLWQLGLAGRREVGARALQ

```

```

HNLGGAIVVTLVAMFPGAASDGVAVPLSAVDGFKHLVKEIEKKLOEEO
PVKRTGVAPEKIKGPGKERTAVVDVKNKGSAVANSDKRADCTIMADTDLALM
TGKMPQVAFRQKIKISGNGMAKNLQLOPKALU"
BASE COUNT 715 a 463 c 608 g 657 t
ORIGIN
Query Match 42.8%; Score 425.6; DB 5; Length 2443;
Best Local Similarity 73.3%; Pred. No. 9.6e-96;
Matches 557; Conservative 0; Mismatches 201; Indels 2; Gaps 1;
QY 6 CCAAGTGTCTCTCAGATGTTGGGTATGCTGAAAAAGAACATATGAAAAATATGAAACA 65
DB 480 CCAATTAACACCGCATGTTTGTCAATGCTGGCAAGAACAATATGAGAAATATACGGGACA 539
QY 66 AAAATTGAACACTTTGCAAAAAATTTGATGGAATAATCATAAACATTCAGTTAATACCGG 125
DB 540 AATCCAGATGACTTTGCAAAAATTCATGGAAGAAATCACAGCATTTCAACCAACACCC 599
QY 126 TATTCGCCAGTTCCAGATGATACAGTTAGTGAAGTATGATGATGATGATGATGATGATG 185
DB 600 TATTCGCCAGTTCCAGATGATACAGTTAGTGAAGTATGATGATGATGATGATGATGATG 659
QY 186 GATTTTGTGACTATGCTTACCAATGTTGTCCCATGATGATGATGATGATGATGATGATG 245
DB 660 GATTTTGTGACTATGCTTACCAATGTTGTCCCATGATGATGATGATGATGATGATGATG 719
QY 246 GCCAGTGAAGCATTTGTACAGAAATATGCGCTGCAATCCAAAGCTGTGAAATTTTGCA 305
DB 720 GCTAGTGAAGCATTTGTGTGAAGGATATGATGATGATGATGATGATGATGATGATGATG 779
QY 306 CAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 365
DB 780 CAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 839
QY 366 GCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 425
DB 840 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 899
QY 426 AATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 485
DB 900 ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959
QY 486 GAAGCACTGGGACTCTGTCAGAAAGACAAGGATGCAAGCTGTGATGATGATGATGATGAT 545
DB 960 GAAGCACTGGGACTCTGTCAGAAAGACAAGGATGCAAGCTGTGATGATGATGATGATGAT 1019
QY 546 ACATATGAGAGAAAGTGGGTCTAATAATCTTATGATGATGATGATGATGATGATGATGATG 605
DB 1020 ACCTAGCGGGGAAAGTGGGTCTAATAATCTTATGATGATGATGATGATGATGATGATGATG 1079
QY 606 CTAGCGGCTACAGGTCTTGTCTCAGTGTGCAAGTCTCTGTGCAAGTCTGAGAGGGAAGCC 665
DB 1080 CTYGTGTCCACAGGCTGTGCGGAGTGTGCAAGTCTCTGTGCAAGTCTGAGAGGGAAGCC 1139
QY 666 GGAAGAAGAGCAAAAGTTCTGTGTCGAAAGTGTGCTGTGCAAGTCTGAGAGGGAAGCC 725
DB 1140 GG--GAGGCGGGAAGTGGGCGGCGCAAGCGGCGGCTGTGCAAGTCTGAGAGGGAAGCC 1197
QY 726 GAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 765
DB 1198 GGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1237

```

Search completed: November 27, 2003, 10:02:39  
 Job time : 4031.04 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:55:53 ; Search time 2679.01 Seconds  
(without alignments)  
9017.749 Million cell updates/sec

Title: US-09-835-992a-21

Perfect score: 994  
Sequence: 1 ctccaccgctgcctccacg.....ccaccctgggtgtgtggat 994

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estcov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rnd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	688.2	69.2	817	9	AU121360 AU121360
2	655.4	65.9	828	14	CD104768 AGENCOURT
3	607.4	61.1	609	14	CB147601 K-EST0203
4	588	59.2	600	10	BE887512 BE887512 601508161

Result No.	Score	Query Match	Length	ID	Description
5	576.6	58.0	1969	11	BC029332 Homo sapi
6	576.6	58.0	1974	11	AK004860 Mus muscu
7	572.6	57.6	926	10	BG289834 602384738
8	557.4	56.1	943	10	BG105684 602312282
9	532.4	53.6	668	10	AV653303 AV653303
10	514	51.7	757	14	CB951253 AGENCOURT
11	512	51.5	816	14	CB950754 AGENCOURT
12	508.8	51.2	964	9	AI525668 PT1.3.04
13	501.4	50.4	583	9	AI694422 DKFZP0313F
14	495.4	49.8	577	9	AL694509 DKFZP0313G
15	470.2	47.3	1051	12	BM552173 AGENCOURT
16	469.2	47.2	525	10	BE971366 601651537
17	457.2	46.0	763	14	CB316398 AGENCOURT
18	443.6	44.6	599	14	CB162403 K-EST0222
19	439.6	44.2	625	14	CB153363 K-EST0210
20	430.8	43.3	731	14	CB947142 AGENCOURT
21	419.6	40.2	849	10	BG432823 602496031
22	407	40.9	1039	9	AL551168 602496031
23	404	40.6	532	9	AI116787 602496031
24	386.6	38.9	545	14	R84426 602496031
25	370.2	37.2	921	13	BX433336 602496031
26	368.4	37.1	805	13	BU962149 AGENCOURT
27	362.2	36.4	1475	10	BF180638 601808747
28	356.6	36.1	526	14	CB158882 K-EST0218
29	356.6	35.9	744	14	CB948021 AGENCOURT
30	355.2	35.7	456	14	F14816 602496031
31	349.2	35.1	532	9	AA105592 602496031
32	341	34.3	669	12	BI147497 602496031
33	332.8	33.5	878	10	BF696675 602127088
34	324.4	32.6	964	10	BG289606 602381573
35	323.6	32.6	678	10	BE911641 601663325
36	321.6	32.4	757	10	BG718743 602666877
37	305.4	30.7	764	14	CB947576 AGENCOURT
38	304.4	30.6	426	14	CB286997 602496031
39	303.4	30.5	461	14	CB286996 602496031
40	302.6	30.4	321	9	AA343982 602496031
41	296.4	29.8	721	14	CD241357 AGENCOURT
42	294.4	29.6	656	12	BI145089 602909165
43	293	29.5	908	13	BQ877242 AGENCOURT
44	290.6	29.2	805	12	BG777467 602664789
45	279.4	28.1	410	9	AW511640 xub0a08.x

## ALIGNMENTS

RESULT 1  
AU121360  
LOCUS  
DEFINITION AU121360 HEMBB1 Homo sapiens CDNA clone HEMBB1002667 5', mRNA  
sequence.  
ACCESSION AU121360  
VERSION AU121360.1 GI:10936595  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,  
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and  
Isogai,T.  
HRI human CDNA project  
Unpublished  
JOURNAL  
CONTACT: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3875  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human CDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; CDNA library construction; Department of

Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute

# FEATURES

source

1. 817  
 /organism="Homo sapiens"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:9606"  
 /clone="HEMB1002667"  
 /issue\_type="whole embryo, mainly body"  
 /dev\_stage="embryo, 10 weeks"  
 /clone\_1b="HEMB1"  
 /note="Vector: pME18SFL3"

BASE COUNT 255 a 141 c 201 g 217 t 3 others

ORIGIN

Query Match 69.2%; Score 688.2; DB 9; Length 817;  
 Best Local Similarity 93.4%; Pred. No. 2.8e-131;  
 Matches 756; Conservative 0; Mismatches 45; Indels 8; Gaps 4;

```

QY 75 CACTTGGCAAAATGATGAGAAATTCATTAACATTCAGTTAATACCCGTAATCCCG 134
DB 1 CACTTGGCAAAATGATGAGAAATTCATTAACATTCAGTTAATACCCGTAATCCCG 60
QY 135 TTCCAAGATGATACAGTTAGATGAGATGAGATGAGATGAGATGAGATGAGATG 194
DB 61 TTCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
QY 195 ACTATCTTCAATGTTGTTCCCACTTCAATGTTGTTCCCACTTCAATGTTGTT 254
DB 121 ACTATCTTCAATGTTGTTCCCACTTCAATGTTGTTCCCACTTCAATGTTGTT 180
QY 255 GGATTTGTACAGAAATGAGTCCGCAATCCAAAGCTGAGAAATTTGGCACAAGAA 314
DB 181 GGATTTGTACAGAAATGAGTCCGCAATCCAAAGCTGAGAAATTTGGCACAAGAA 240
QY 315 ATGACTGATTTGCCAAGCTCGTTTGAAGAAAGAAAGATTAATTAATGTTGGCTT 374
DB 241 ATGACTGATTTGCCAAGCTCGTTTGAAGAAAGAAAGATTAATTAATGTTGGCTT 300
QY 375 ATGAGTAAGAAAGCTGCAAGAAATGCTATGAGAAATGCTGCTGACACCAATGAT 434
DB 301 ATGAGTAAGAAAGCTGCAAGAAATGCTATGAGAAATGCTGCTGACACCAATGAT 360
QY 435 GAGCTATGAGAACTTCAAGATGCTTCTTCAAGAACTTCTTCAAGAACTTCTTCA 494
DB 361 GAGCTATGAGAACTTCAAGATGCTTCTTCAAGAACTTCTTCAAGAACTTCTTCA 420
QY 495 GGAAGCTGTCAGAAAGCAAGGTCAGAGCTGTTGATAGAGAGATTAATCATATG 554
DB 421 GGAAGCTGTCAGAAAGCAAGGTCAGAGCTGTTGATAGAGAGATTAATCATATG 480
QY 555 GGAAGGTTGGCTATTAATCTTCTAGTGTGAGTGTATTCAGAGGACACCCACTAG 614
DB 481 GGAAGGTTGGCTATTAATCTTCTAGTGTGAGTGTATTCAGAGGACACCCACTAG 540
QY 615 AAGAGCTTGTCTCAGTGTGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 674
DB 541 AAGAGCTTGTCTCAGTGTGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 598
QY 675 GCAAGATTTCTGCTGCAAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 734
DB 599 GCAAGATTTCTGCTGCAAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658
QY 735 TTGTAACTCTTCAAGATGAGTGGGTTTCCCGAAGCCGCAAGTCTCTTTTGAAC 794
DB 659 TTGTAACTCTTCAAGATGAGTGGGTTTCCCGAAGCCGCAAGTCTCTTTTGAAC 716
QY 795 AAAATGGAAGCNGTTCCAAAGCTGCAAGTATGATGATGATGATGATGATGATG 851
DB 717 AAAATGGAAGCNGTTCCAAAGCTGCAAGTATGATGATGATGATGATGATGATG 775
QY 852 GTTTTAAAGGAGATGAGAGAAACNT 880

```

DB 776 GTTTTAAAGGAGATGAGAGAAACNT 804

RESULT 2  
 CD104768 828 bp mRNA linear EST 15-MAY-2003  
 LOCUS  
 DEFINITION AGENCOURT 14007012 NIH MGC 186 Homo sapiens cDNA clone  
 IMAGE:30372857 5', mRNA sequence.  
 ACCESSION CD104768  
 VERSION CD104768.1 GI:30757942  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-rcmail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDCM165 row: 1 column: 18  
 High quality sequence stop: 463.

## FEATURES

source

1. 828  
 /organism="Homo sapiens"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30372857"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_1b="NIH MGC 186"

/note="Organ: Pooled-Skin; Vector: pDNR-LIB; Site 1: SfiI  
 (ggccatctggcc); Site 2: SfiI (ggccctcgccg); Library is  
 oligo-dT primed and directionally cloned. cDNA was  
 prepared from a pooled samples of tissues from skin,  
 meninges, dura mater, pia mater and choroid plexus. 5'  
 and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CAGGCGCATTTATGCG-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCGCGAGCGCGGACATG-dT(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
 Library"

BASE COUNT 261 a 156 c 195 g 216 t

ORIGIN

Query Match 65.9%; Score 655.4; DB 14; Length 828;  
 Best Local Similarity 94.9%; Pred. No. 1.5e-124;  
 Matches 719; Conservative 0; Mismatches 34; Indels 5; Gaps 4;

```

QY 60 GGAACAAAATTTGAACATTTGCAAAATTTGATGAGAAATTCATTAACATTCAGTTAT 119
DB 3 GGAACAAAATTTGAACATTTGCAAAATTTGATGAGAAATTCATTAACATTCAGTTAT 62
QY 120 AACCGATTTCCAGTTCCAGATGATGATGATGATGATGATGATGATGATGATG 179
DB 63 AACCGATTTCCAGTTCCAGATGATGATGATGATGATGATGATGATGATGATG 122
QY 180 GTTTTGAATTTTGAATCTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 239
DB 123 GTTTTGAATTTTGAATCTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 182
QY 240 ATTTTGGCAGTGAAGATTTGTACAGAAATGATGAGCTGCAATCCAAAGCTGTGAAATT 299

```

```

Db      183 ATTTGGCCAGTGAAGCATTTGTACAGAGATAGTGGCTGCATCCAAAGCTGTGAAATT 242
Oy      300 TTGGCACAAGAAATATGATGATCTGATTTGCCAAGCTGTTGAAGAAAAGCATTTATTA 359
Db      243 TTGGCACAAGAAATATGATGATCTGATTTGCCAAGCTGTTGAAGAAAAGCATTTATTA 302
Oy      360 ATGTTGGCTTTGATATGATGATTAAGAGCTGCAGAAAATGCTATGAGAAATCTGGCCCTG 419
Db      303 ATGTTGGCTTTGATATGATGATTAAGAGCTGCAGAAAATGCTATGAGAAATCTGGCCCTG 362
Oy      420 ACACCAAAATGATATGATGATTAAGAGCTGCAGAAAATGCTATGAGAAATCTGGCCCTG 479
Db      363 ACACCAAAATGATATGATGATTAAGAGCTGCAGAAAATGCTATGAGAAATCTGGCCCTG 422
Oy      480 ACTTATGAAAGCACTGGGACTCTGTCCAGAAAGCAAGGTCGAAACCTGTTTATGAGAGA 539
Db      423 ACTTATGAAAGCACTGGGACTCTGTCCAGAAAGCAAGGTCGAAACCTGTTTATGAGAGA 482
Oy      540 GATTAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGTGAGACTGATTTCAAAGGA 599
Db      483 GATTAATACATATGAGAGAGAAAGTGGTCAATTAATCTAGTGTGAGACTGATTTCAAAGGA 542
Oy      600 CACCCACTAGAGCGCTACAGAGTCTTGCTCAGTGTGAGAACTCTGCTGCGAGCTGAGAGG 659
Db      543 CACCCACTAGAGCGCTACAGAGTCTTGCTCAGTGTGAGAACTCTGCTGCGAGCTGAGAGG 602
Oy      660 GAAGCCGAGAAAGAGGCAAAAGTCTCTGCTGCAAGAGTGTG-CTCTGCGATTAATTATTA 718
Db      603 GAAGCCGAGAAAGAGGCAAAAGTCTCTGCTGCAAGAGTGTG-CTCTGCGATTAATTATTA 662
Oy      719 ATTGAGAGAA--CTGTGTTGTATACA-CTCTCAAGATGGGGTTTTCCCGG-AGCGCGC 774
Db      663 ATTGAGAGAGAGCTGTGTTGTATACA-CTCTCAAGATGGGGTTTTCCCGGAGAAAGG 722
Oy      775 AGTTCCTTTTGAAGCTCATCAAAATTTGAAGCCNGTTCC 812
Db      723 AGTTCCTTTTGAAGCTCATCAAAATTTGAAGCTGTTTC 760

RESULT 3
LOCUS   CBI47601                      609 bp    mRNA    linear    EST 29-JAN-2003
DEFINITION K-EST0203627 L11SNUS3481 Homo sapiens cDNA clone L11SNUS3481-29-G02
5' mRNA sequence.
ACCESSION CBI47601
VERSION   CBI47601.1 GI:28129068
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS  Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE     21C Frontier Korean EST Project 2001
JOURNAL   Unpublished
COMMENT   Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boseun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 29 row: G column: 02
High quality sequence stop: 609.
location/Qualifiers
1..609
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L11SNUS3481-29-G02"
/sex="M"

```

```

/tissue_type="Liver"
/cell_type="Polygonal"
/cell_line="SNU-354"
/lab_host="Top10F"
/clone_id="L11SNUS3481"
/notes="Organ: Liver; Vector: pcns-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then dephosphorylated
with tobacco acid pyrophosphatase (TAP). The dephosphorylated
intact mRNA was ligated with RNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and Nid14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transfection of the
remaining DNA into competent cells E. coli Top10F with
electroporation method."
BASE COUNT 203 a 105 c 137 g 164 t
ORIGIN
Query Match 61.1%; Score 607.4; DB 14; Length 609;
Best Local Similarity 99.8%; Pred. No. 11e-114;
Matches 608; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy      45 CATATGAGAAATATGAGAAATTTGAACATTTGCAAAATTTGATGAGAAATCAT 104
Db      1 CATATGAGAAATATGAGAAATTTGAACATTTGCAAAATTTGATGAGAAATCAT 60
Oy      105 AAACATTCAGTTAATTAACCCGTTATCCAGTTCCAAATGATTAAGTTAGATGAAGT 164
Db      61 AAACATTCAGTTAATTAACCCGTTATCCAGTTCCAAATGATTAAGTTAGATGAAGT 120
Oy      165 ATGCGATCTAAAGAGTTTGTGATTTTGTGACTATCTTACAAATGTTGCCCATTCAGAT 224
Db      121 ATGCGATCTAAAGAGTTTGTGATTTTGTGACTATCTTACAAATGTTGCCCATTCAGAT 180
Oy      225 GGTGCTGAGCAGCAATTTTGGCCAGTGAAGCATTTGTACAGAAATATGAGCTGCAATCC 284
Db      181 GGTGCTGAGCAGCAATTTTGGCCAGTGAAGCATTTGTACAGAAATATGAGCTGCAATCC 240
Oy      285 AAAGCTGTGAATTTTGGCACAAGAAATGATGACTGATTTGCCAAGCTGTTGAAGAA 344
Db      241 AAAGCTGTGAATTTTGGCACAAGAAATGATGACTGATTTGCCAAGCTGTTGAAGAA 300
Oy      345 AAAGCATTTTAAATGTTGCTGTTGATATGATTAAGAGCTGCAGAAATATGCTAT 404
Db      301 AAAGCATTTTAAATGTTGCTGTTGATATGATTAAGAGCTGCAGAAATATGCTAT 360
Oy      405 GAGAAATCTGGCTGACACCAATATGATTAAGAGCTGCAAGATTTGCTTTTCT 464
Db      361 GAGAAATCTGGCTGACACCAATATGATTAAGAGCTGCAAGATTTGCTTTTCT 420
Oy      465 ACCAAGCAATCTCTTACTTATGAGAGCTGGAGCTCTGTCCAGAAAGCAAGGTCACAG 524
Db      421 ACCAAGCAATCTCTTACTTATGAGAGCTGGAGCTCTGTCCAGAAAGCAAGGTCACAG 480

```

Qy	525	CTGGTGTATTAGAGGAGTAATAATACATATATGAGGAAGTGGGTCATCAAATCCATAGTGTGA	584
Db	481	CTGGTGTATTAGAGGAGTAATAATACATATATGAGGAAGTGGGTCATCAAATCCATAGTGTGA	540
Qy	585	CTGATTTCAAAGGACACCCACTAGACGCTCTTGGCTCAGTGTGCAGAACTCTGC	644
Db	541	CTGATTTCAAAGGACACCCACTAGACGCTCTTGGCTCAGTGTGCAGAACTCTGC	600
Qy	645	TGGCAGCTG	653
Db	601	TGGCAGCTG	609
RESULT 4			
BE887512			
LOCUS	BE887512	600 bp	mRNA linear EST 20-OCT-2000
DEFINITION	601508161F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909614 5',		
ACCESSION	BE887512		
VERSION	BE887512.1	GI:10342878	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 600)		
TITLE	NIH-MGC htcp://mhc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLN at:		
	htcp://image.lnl.gov		
	Plate: LLM9723 row: j column: 15		
	High quality sequence stop: 598.		
FEATURES			
source	location/Qualifiers		
	1..600		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3909614"		
	/tissue_type="leiomyosarcoma"		
	/lab_host="RDH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_71"		
	/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;		
	Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.		
	Average insert size 2.1 kb.		
BASE COUNT	204 a 101 c 135 g 160 t		
ORIGIN			
Query Match	59.2%; Score 588; DB 10; Length 600;		
Best Local Similarity	99.8%; Pred. No. 1.le-110;		
Matches	599; Conservative. 0; Mismatches 0; Indels 1; Gaps 1;		
Qy	36	GGAAAGAACATATGGAATAATATGGAACAAAATTGACACTTTGCAAAAATTGATGG	95
Db	1	GGAAAGAACATATGGAATAATATGGAACAAAATTGACACTTTGCAAAAATTGATGG	60
Qy	96	AAAATCATTAACACATTGATTATATTAACCCGATTCGCCAAGATGAATACAGTTTA	155
Db	61	AAAATCATTAACACATTGATTATATTAACCCGATTCGCCAAGATGAATACAGTTTA	120
Qy	156	GATGAAGTGAATGCGATCTTAAGAAGTTTGTATTTTGTACTATCTTACATGTTGTCCC	215
Db	121	GATGAAGTGAATGCGATCTTAAGAAGTTTGTATTTTGTACTATCTTACATGTTGTCCC	180
Qy	216	ACTTCAGATGCTGTCAGCAGCAATTTGGCCAGTGAAGCATTTGTACAGAGTATGCC	275

Db	Accession	Definition	Version	Keywords	Source	Organism	Reference	Title	Journal	Remark	Comment
Db	181	ACCTCAGATGATGCTGCAGCAGCAATTTTGGCAGTGAACCATTTGTACAGAAATATAGC	240								
Db	276	CTGCATATCCAAAGCTGTGGAAATTTTGGCAACAAGAAATGATGACTGATTTTGGCAAGCTCG	335								
Db	241	CTGCATATCCAAAGCTGTGGAAATTTTGGCAACAAGAAATGATGACTGATTTTGGCAAGCTCG	300								
Db	336	TTTGAACAAAAAAGCATTTATTAATATGTTGCTTTGATATGATGAAGAAGCTGCAGA	395								
Db	301	TTTGAACAAAAAAGCATTTATTAATATGTTGCTTTGATATGATGAAGAAGCTGCAGA	360								
Db	396	AAATGCTATGAGAATCTGGCCTGCACACAAATGATATTGACGTAAATGAACATTCACGAT	455								
Db	361	AAATGCTATGAGAATCTGGCCTGCACACAAATGATATTGACGTAAATGAACATTCACGAT	420								
Db	456	TGCTTTTCTACCAACGACTCTTACTTATGAAAGCACTGGGACTCTGTCCAGAAAGCA	515								
Db	421	TGCTTTTCTACCAACGACTCTTACTTATGAAAGCACTGGGACTCTGTCCAGAAAGCA	480								
Db	516	GGTGAACGCTGGTGTAGTAGAGATATATACATATGAGGAAGAGGGGATTAATTC	574								
Db	481	GGTGAACGCTGGTGTAGTAGAGAGATATATACATATGAGGAAGAGGGGATTAATTC	540								
Db	575	TAGTGTGACATGATTTCAAAAGGACACCCACTAGGCGCTACAGCTTTCCTCAGTGTGC	634								
Db	541	TAGTGTGACATGATTTCAAAAGGACACCCACTAGGCGCTACAGCTTTCCTCAGTGTGC	600								
RESULT 5											
LOCUS	BC029332	1969 bp	mRNA	linear	HTC	01-MAY-2002					
DEFINITION	BC029332	Homo sapiens, clone IMAGE:5109488,	mRNA.								
VERSION	BC029332.1	GI:20379648									
KEYWORDS	HTC.										
SOURCE	Homo sapiens (human)										
ORGANISM	Homo sapiens										
REFERENCE	Eukaryote, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;										
AUTHORS	Mammalia, Eutheria, Primates, Catarrhini, Homnidae, Homo.										
TITLE	1 (bases 1 to 1969)										
JOURNAL	Strausberg, R.										
	Direct Submission										
	Submitted (01-MAY-2002) National Institutes of Health, Mammalian										
	Gene Collection (MGC), Cancer Genomics Office, National Cancer										
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,										
	USA										
	NIH-MGC Project URL: http://mgc.nci.nih.gov										
	Contact: MGC help desk										
	Email: cgabs-remail.nih.gov										
	Tissue Procurement: ATCC										
	CNA Library Preparation: Life Technologies, Inc.										
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)										
	DNA Sequencing by: Baylor College of Medicine Human Genome										
	Sequencing Center										
	Center code: BCM-HGSC										
	Web site: http://www.hgsc.bcm.tmc.edu/cdna/										
	Contact: amg@bcm.tmc.edu										
	Guarantees, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,										
	Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,										
	Richards, S., Gibbs, R.A.										
	Clone distribution: MGC clone distribution information can be found										
	through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov										









BASE COUNT	207 a	116 c	176 g	168 t	1 others
ORIGIN					
Query Match		53.6%	Score 532.4	DB 9	Length 668
Best Local Similarity		92.7%	Pred. No. 3e-99		
Matches 608	Conservative	0	Mismatches 38	Indels 10	Gaps 5
Qy	229	CTGCAGCAGCAATTTTGGCCAGTGAAGACATTTGTACAGAAAGATGAGCCCTGCATCCAAAG	288		
Db	1	CTGCAGCAGCAATTTTGGCCAGTGAAGACATTTGTACAGAAAGATGAGCCCTGCATCCAAAG	60		
Qy	289	CTGTGAAATTTTGGCACACAGAAATGATGACTGATTTGCCAAGCTCGTTTGAAGAAAAA	348		
Db	61	CTGTGAAATTTTGGCACACAGAAATGATGACTGATTTGCCAAGCTCGTTTGAAGAAAAA	120		
Qy	349	GCATTATTTAAATGGTTGGCTTTTGATATGAGTAAAGAAAGCTGCAGAAATATCTATGACA	408		
Db	121	GCATTATTTAAATGGTTGGCTTTTGATATGAGTAAAGAAAGCTGCAGAAATATCTATGACA	180		
Qy	409	AATCTGGCGCTGACACCAAATGATATTGACGTATGAACTTCACAGTTGCTTTCTACCA	468		
Db	181	AATCTGGCGCTGACACCAAATGATATTGACGTATGAACTTCACAGTTGCTTTCTACCA	240		
Qy	469	ACGAAGCTCTTACTTATGAAAGCACTGGGACTCTGTCCAGAAAGACAAGAGTCGACAGCTGG	528		
Db	241	ACGAAGCTCTTACTTATGAAAGCACTGGGACTCTGTCCAGAAAGACAAGAGTCGACAGCTGG	300		
Qy	529	TTGATATAGAGGAGATTAATTCATATGAGAAAGTGGGTATTAATCTTACTGTGGACTCTCA	588		

QY 589 TTTCAAAGGACACCCACTAGGCGCTACAGGTCTTGCTAGTGTGCAGAACTCTGTGGC 648

[illegible]

b6  
421 A C T G A A C C C C C A A - G A G C A A G T T C C T G T C A A A G T G C C T C T G C A G C A 478

709 TAATTTCATTTGAGGAAGTGTGTTGTACACTCTACAGATGGGGTTTCCCGAA 768

Db 479 TAATTAGCATTGGAGGAGCTGTGGTTGTACACTCTACAGATGGG--TTTTCGGAA 536

769 GCCGCCAGTTCCTTTTAGAACTCATCAAATTTGAAGCCNGTTCCAACCAAGCTTGCAAG 828

Db 537 GCCGCCAGTT-CTTTAGACTCATCANATTGAA---GCTGTTCCAACCAGCTCTGCAAG 592

829 TNATNGGTTTAANGNAAATCTNGCTTTTAAAGGNGGATTGAGAAGGAAACNTNAAA 884

DB 593 TGA TGGATTTAAGGCATACTGTGTTTAA--GGAGATTGAGAGAGAAATCTGAAAGA 646

## RESULT 10

LOCUS	CB951253	757 bp	mRNA	linear	EST 29-APR-2003
-------	----------	--------	------	--------	-----------------

IMAGE:30316161 5', mRNA sequence.

VERSION CB951253.1 GI:30206290  
ZEVWDDDC FCM

SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>.

**JOURNAL Unpublished**

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

CDNA Library Preparation: Michael Brownstein Laboratory  
CDNA Library: Michael Brownstein Laboratory (IMB)  
CDNA Library: Michael Brownstein Laboratory (IMB)

DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: NDCM107 row: b column: 10  
 High quality sequence stop: 551.  
 Location/Qualifiers

## FEATURES

SOURCE

1.757  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30322082"  
 /lab\_host="DH10B (T1-phage-resistant)"  
 /clone\_lib="NIH MGC 177"  
 /note="Organ: Liver; Vector: pDNR-LIB; Site 1: SfiI  
 (ggccatcggcc); Site 2: SfiI (ggcgccctggcc); cDNA made  
 by oligo-dt priming and directionally cloned. 5' and 3'  
 adaptors were used in cloning as follows:  
 5'-AAGCAGTGGTATCAACGGGCGGCGGCGG-3' and  
 5'-ATTCTAGAGCGGCGGCGGCGGCGGCGG-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the 0.5 kb  
 size fraction. Library created in the laboratory of M.  
 Brownstein (NIH, NIH). Note: this is a NIH\_MGC Library."  
 BASE COUNT 212 a 174 c 210 g 160 t 1 others  
 ORIGIN

Query Match 51.7%; Score 514; DB 14; Length 757;  
 Best Local Similarity 83.8%; Pred. No. 1.8e-95;  
 Matches 605; Conservative 0; Mismatches 112; Indels 5; Gaps 2;

50 GGAATAATATGGAACAAATATGAACTTTGCAAAAATTTGATGAAAAATCATTAACA 109  
 3 GGAATAATATGGAACAAATATGAACTTTGCAAAAATTTGATGAAAAATCATTAACA 62  
 110 TTCAGTATAATACCGTATTCAGTATTCAGATGTAATGATGATGATGATGATGATG 169  
 63 CTCAGTATAATACCGTATTCAGTATTCAGATGTAATGATGATGATGATGATGATG 122  
 170 ATCTAAGAAGTTTGTGATTTTGTGATGATGATGATGATGATGATGATGATGATG 229  
 123 ATCAAAACAGTTTTCGATTTTCTGATGATGATGATGATGATGATGATGATGATG 182  
 220 TGCAGCAGCAATTTTGGCGAGTGAAGCATTTGTACAGAAGTATGGCTTCATCCAAAG 289  
 183 CGCTCCCGGATTCGTCCAGCAGAGAGTTTGTGACAGTACGGCTCCATCCAAAGC 242  
 290 TGTGAAATTTTGGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATG 349  
 243 GGTGAGATGTTGGCCCGAGAGATGATGATGATGATGATGATGATGATGATGATG 302  
 350 CATTAATAAATGTTGGCTTGTATGATGATGATGATGATGATGATGATGATGATG 409  
 303 TATTAATAAAGTGTGGTGTATGATGATGATGATGATGATGATGATGATGATGATG 362  
 410 ATCTGCTGACACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 469  
 363 GTCCGCGCTGACACCAATGATGATGATGATGATGATGATGATGATGATGATG 422  
 470 CGAATCTCTTATTAAGAAGCATGGAATCTGTCCAGAAAGCAAGGTCAGCTGGT 529  
 423 TGAATCTCTTATTAAGAAGCATGGAATCTGTCCAGAAAGCAAGGTCAGCTGGT 482  
 530 TGATAGAGAGATAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGTGAGTAT 589  
 483 GGACAGAGAGAGATAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGTGAGTAT 542  
 550 TTCAAAAGGACACCACTAGGCGCTACAGGTCTTGTCTCAGTGTGAGAACTTGTCTGCA 649  
 543 CTCAAAGGACACCACTAGGCGCTACAGGTCTTGTCTCAGTGTGAGAACTTGTCTGCA 602  
 650 GGTGAGAGAGAGCGGAAAGAGCAAGTTCTGTCGCAAGGTGCTGTGNGCAT 709

Db 603 GCTGAGAGCGGACCGGAAA--GAGCAGGTTCCCGGCGGCAAAAGTGGCTTCGACAGC 660  
 Qy 710 AATTATGACATTTGAGAGAACTGTGTGTAACACTTACA--AGATGGGTTTCCCGG 766  
 Db 661 AATCTAGGCTCTGGAGAGAGTGTGTGTCACCCCTCTACANAGATGGGGTTTCCCGG 720  
 Qy 767 AA 768  
 Db 721 AA 722

## FEATURES

SOURCE

RESULT 11  
 CB950754 866 bp mRNA linear EST 29-APR-2003  
 LOCUS CB950754  
 DEFINITION AGENCOURT 13777269 NIH\_MGC 177 Mus musculus cDNA clone  
 IMAGE:30322082 5', mRNA sequence.  
 ACCESSION CB950754  
 VERSION CB950754.1 GI:30205281  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 866)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Stransberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Michael Brownstein Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDCM122 row: i column: 03  
 High quality sequence stop: 481.  
 Location/Qualifiers

## FEATURES

SOURCE

1.866  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30322082"  
 /lab\_host="DH10B (T1-phage-resistant)"  
 /clone\_lib="NIH MGC 177"  
 /note="Organ: Liver; Vector: pDNR-LIB; Site 1: SfiI  
 (ggccatcggcc); Site 2: SfiI (ggcgccctggcc); cDNA made  
 by oligo-dt priming and directionally cloned. 5' and 3'  
 adaptors were used in cloning as follows:  
 5'-AAGCAGTGGTATCAACGGGCGGCGGCGGCGG-3' and  
 5'-ATTCTAGAGCGGCGGCGGCGGCGGCGGCGG-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the 0.5 kb  
 size fraction. Library created in the laboratory of M.  
 Brownstein (NIH, NIH). Note: this is a NIH\_MGC Library."  
 BASE COUNT 240 a 200 c 242 g 183 t 1 others  
 ORIGIN

Query Match 51.5%; Score 512; DB 14; Length 866;  
 Best Local Similarity 82.8%; Pred. No. 4.6e-95;  
 Matches 607; Conservative 0; Mismatches 123; Indels 3; Gaps 2;

81 GCAAAATTTGATGAGAAATCATTAACATTCAGTTAATACCCGATATCCAGTTCAA 140  
 3 GCAAAATTTGATGAGAAATCATTAACATTCAGTTAATACCCGATATCCAGTTCAA 62  
 141 GATGAATACAGTTTATGATGAAGTATGATGATGATGATGATGATGATGATGATG 200  
 63 GATGAATACAGTTTATGATGAAGTATGATGATGATGATGATGATGATGATGATG 122  
 201 TTCAATGTTGCCACTTCAGATGATGATGATGATGATGATGATGATGATGATGATG 260

Db	123	TTGCAATCTGCGCCACCTCAGATGTGTCGCCGTGCGGAGTTCTGTCCAGGAGAGATT	182
Qy	261	GTACGAAGATATGGCTCTGCAATCCAAAGCTGTGAAATTTTGGCACAAGAAATGATGACT	320
Db	183	GTGCGAGCGATACGGCTTCAGTCCAAACCGGTGGAGATTGTGGCCACAGAAATGATGACT	242
Qy	321	GATTGGCCAAAGCTGTTTGAAGAAAAAGCATTTATTAATGATGGTGGCTTTGATATGAGT	380
Db	243	GACTTACCCAGTACCTTTGAAAGAGAAAGTATTTATTTAAAGTGTGGCTTATGATATGAGT	302
Qy	381	AAAGAGCTGCAGAAGAAAATGCTATGAGAAATCTGGCTTGACACCAATATGATTTGACGTA	440
Db	303	AAAGAGCTGCACAGAGATGCTATGAGAAAGTCCGGCTGCACACCAAGATGTGACGTG	362
Qy	441	ATGAGACTTCACAGATTGCTTTTCTACCAACAAGACTCTTACTTATGAAAGCACTGGAGTCT	500
Db	363	ATAGAGCTTCACGATTTGCTTCTCTGTCAATGAACATCATCTTACGAAAGCACTGGGGCTC	422
Qy	501	TGTCCAGAAGGACAAAGTGCACACGCTGGTTGATGAGAGATATGACATATGAGGAAAG	560
Db	423	TGTCCAGAAGGACAAAGTGCACACCTGTGTGACAGAGGGGACAACTTACGAGGAAAG	482
Qy	561	TGGGTGATATAATCCTAGTGTGAGACTGATTTCAAAGGGACACCACTAGGCGCTACAGT	620
Db	483	TGGGTGATCAACCTTAGTGTGAGAGCTCATCTCAAAGGACACCACTGGGCGCAACAGT	542
Qy	621	CTTGCTCAGTGTGCAGAACTGTGCTGCGAGCTGAGAGGGGAAACCGGAAAAAGGCAAG	680
Db	543	CTGGCTCAGTGTGCGGAGACTCTGTGCTGAGCTGAGAGCGGAAACCGGAAA--GAGGCAAG	600
Qy	681	TTCTGTGTGCAGAAAGTGTGCTGTGCNGCATTAATTANGATTTGAGAGAACTGTGTGTGTA	740
Db	601	TTCCCGGGGCAAAATGTGCTCTGCAGCACAATCTAGGCTGTGAGAGAACTGTGTGTGTC	660
Qy	741	CACCTTACAAATGTGGGGTTTTCCGGAGAGCGCCAGTCTCTTTAGAATCATCAAAATT	800
Db	661	ACCTTACAGATGGGTTTTTCCC-GAAGCTGCCAGTCCCTTCAGAGACGACACAGATT	719
Qy	801	GAAGCCNGTTCCA 813	
Db	720	TCAGCTGAGCTCCA 732	
RESULT 12			
LOCUS	A1525668	914 bp	mRNA linear EST 09-AUG-1999
DEFINITION	PT1.3_04 D09.r tumo01 Homo sapiens cDNA 5', mRNA sequence.		
ACCESSION	A1525668		
VERSION	A1525668.1	GI:4439803	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 914) Huang,G.M., Ng,W.L., Farakas,J., He,L., Liang,H.A., Gordon,D., Yu,J. and Hood,L.		
TITLE	Prostate cancer expression profiling by cDNA sequencing analysis		
JOURNAL	Genomics 59 (2), 178-186 (1999)		
MEDLINE	99393982		
PUBMED	10409429		
COMMENT	Contact: Guyang Matthew Huang Leroy Hood University of Washington Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195 Tel: 5106280100 Fax: 5106280108 Email: huanggm@yahoo.com. Location/Qualifiers		
FEATURES	1..914		
SOURCE	/organism="Homo sapiens"		
	/mol_type="mRNA"		

	/db_xref="taxon:9606"	
	/clone_lib="tumor1"	
	/note="Organ: Prostate; Vector: pBluescript; Directional	
	cDNA library was constructed using Lambda ZP II kit	
	(Stratagene). mRNA was extracted from a frozen prostate	
	tumor tissue (Mayo Clinics)."	
BASE COUNT	258 a    174 c    255 g    189 t    38 others	
ORIGIN		
Query Match	51.2%; Score 508.8; DB 9; Length 914;	
Best Local Similarity	86.3%; Pred. No. 2.1e-94;	
Matches	581; Conservative 0; Mismatches 84; Indels 8; Gaps 3;	
Oy	GCCAAGCTCGTTTAAAGAAAAAAGCATTATTAATAATGGTTGCTTGATATGACTAAGA	385
Dd	GCCAAAGCTCGTTGAAGAAAAAAGCATTTATTAATGTGGCTTGTATATGAGTAAGA	78
Oy	AGCTGCAAGAAAATGCTATGAGAAATCGGCTTGACACCATAATGATTTGACGTAAAGA	138
Dd	AGCTGCAAGAAAATGCTATGAGAAATCGGCTTGACACCATAATGATTTGACGTAAAGA	138
Oy	ACTTCACAGATTCCTTTCTAACCAAGAACTCTTACTTATGAAGCACTGGGACTCTGCC	505
Dd	ACTTCACAGATTCCTTTCTAACCAAGAACTCTTACTTATGAAGCACTGGGACTCTGCC	198
Oy	AGAAAGACAAGATGCAACGCTGGTGTATAGAGAGATATACATATGAGAAAGTGGGT	565
Dd	AGAAAGACAAGATGCAACGCTGGTGTATAGAGAGATATACATATGAGAGAAAGTGGGT	258
Oy	CATAAATCTTAGTGTGGAGCTGATTTCAAAGGGGACACCACTAGGCGCTACAGGCTTGC	625
Dd	CATAAATCTTAGTGTGGAGCTGATTTCAAAGGGGACACCACTAGGCGCTACAGGCTTGC	318
Oy	TCACTGTGCAGAACTCTGTGCTGCAGCTGAGAGGGGAAACCGGAAAAGGCAAAATTCT	685
Dd	TCACTGTGCAGAACTCTGTGCTGCAGCTGAGAGGGGAAACCGGAAA--GAGGCAAGTTCT	376
Oy	GGTGCAAAAGGTGGCTCTGCNCGATAATTANGCATTGGAGGAACGTGGTTGTAACCTC	745
Dd	GGTGCAAAAGGTGGCTCTGCNCGATAATTANGCATTGGAGAGAGCTGGTTGTAACCTC	436
Oy	TACAAGATGGGGTTTTTCCGGGAAGCGCCGAGTCCCTTTAGAACTCATCAAAATTGAGC	805
Dd	TACAAGATGGG--TTTTCCGGGAGCCGCCAGTCTTTTGAACCTCTCCAATTGGAGC	494
Oy	CNGTTCACAACAAGCTCTGCAAGTNATNGSTTTAANG---NNAAATCTNGTTTYTAAGG	861
Dd	TGTTCCCAACCCAGCTCTGCAAGTGANGTATTTAAAGGCAAAACCTGTTTTTAAGAAG	554
Oy	NAGATTGAGAAGAAACNTNAAGAGGAANGGGGAACAATTGTGTAAGAAAAATNCG	921
Dd	GGGTTGNNAAGAAACTTTGAAGAGGAANGGGGAACCAATTGTGTAAGAAAAACCGGGN	614
Oy	NGGGAATTTTGGCCCCCTTCAAGGGGAANAANTGCCCCTGGGGGTATAAAGANGCCACCT	981
Dd	GGNAATTTTNNCCCTCCAAAGGNGGAAAGATGGCCCTGGGGGGGAAAAAAAGGCCNCCT	674
Oy	GGGGTGTGGGAT 994	
Dd	GGGGTGTGGGAT 687	
RESULT 13		
AL694422	583 bp	mRNA
LOCUS	DKRZpJ3F1641_r1_313 (synonym: h1cc2)	Homo sapiens CDNA clone
DEFINITION	DKRZpJ3F1641_5' , mRNA sequence.	
ACCESSION	AL694422	
VERSION	AL694422.1 GI:19617879	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

REFERENCE  
1 (bases 1 to 583)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
AUTHORS  
Koehler, K., Beyer, A., Mewes, W., Weill, B. and Wiemann, S.  
TITLE  
EST (Koehler, K., Beyer, A., Mewes, W., Weill, B. and Wiemann, S.)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Koehler K  
MIPS  
Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert.  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by BMFZ (Biomedical Research Center at the Charité,  
Berlin/Germany) within the cDNA sequencing consortium of the German  
Genome Project.  
No. 81 sequence available.  
This clone (DKFZp313f1641) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers

FEATURES  
SOURCE  
1..583  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp313f1641"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_id="313 (synonym: h1cc2)"  
/note="Vector: pTriblex2; Site\_1: SfiI; Site\_2: SfiIb;  
cDNA-collection"

BASE COUNT  
199 a 97 c 121 g 166 t  
ORIGIN

Query Match 50.4%; Score 501.4; DB 9; Length 583;  
Best Local Similarity 99.8%; Pred. No. 7.1e-93;  
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
60 GGAAACAAAATTGAACATCTTGCAAAAATTGGATGGAAAAATCATTAACATTCAGTTAT 119  
81 GGAAACAAAATTGAACATCTTGCAAAAATTGGATGGAAAAATCATTAACATTCAGTTAT 140  
120 AACCGTATTCAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179  
141 AACCGTATTCAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 200  
180 GTTTTGAATTTTGAATCTTACATGTTGTCCCACTTCAGATGATGATGATGATGATGATGAT 239  
201 GTTTTGAATTTTGAATCTTACATGTTGTCCCACTTCAGATGATGATGATGATGATGATGAT 260  
240 ATTTGGCAGTGAAGCATTTGTACAGAAATGATGATGATGATGATGATGATGATGATGATGAT 299  
261 ATTTGGCAGTGAAGCATTTGTACAGAAATGATGATGATGATGATGATGATGATGATGATGAT 320  
300 TTGGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359  
321 TTGGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 380  
360 ATGTTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419  
381 ATGTTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 440  
420 ACACCAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 479  
441 ACACCAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500  
480 ACTTATGAAGCACTGCTGCTCAGAAAGCAAGTGCACACGCTGTTGATAGAGA 539  
501 ACTTATGAAGCACTGCTGCTCAGAAAGCAAGTGCACACGCTGTTGATAGAGA 560  
540 GATATATCATATGAGAGAAAGTG 562  
561 GATATATCATATGAGAGAAAGTG 583

RESULT 14  
AL694509 577 bp mRNA linear EST 21-MAR-2002  
LOCUS  
DEFINITION  
DKFZp313G1541.r1 313 (synonym: h1cc2) Homo sapiens cDNA clone  
AL694509  
ACCESSION  
AL694509  
VERSION  
AL694509.1 GI:19617887  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
1 (bases 1 to 577)  
Koehler, K., Beyer, A., Mewes, W., Weill, B. and Wiemann, S.  
TITLE  
EST (Koehler, K., Beyer, A., Mewes, W., Weill, B. and Wiemann, S.)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Koehler K  
MIPS  
Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany  
This is the 5' sequence of the clone insert.  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by BMFZ (Biomedical Research Center at the Charité,  
Berlin/Germany) within the cDNA sequencing consortium of the German  
Genome Project.  
No. 81 sequence available.  
This clone (DKFZp313G1541) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers

FEATURES  
SOURCE  
1..577  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp313G1541"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_id="313 (synonym: h1cc2)"  
/note="Vector: pTriblex2; Site\_1: SfiI; Site\_2: SfiIb;  
cDNA-collection"

BASE COUNT  
196 a 97 c 119 g 165 t  
ORIGIN

Query Match 49.8%; Score 495.4; DB 9; Length 577;  
Best Local Similarity 99.8%; Pred. No. 1.2e-91;  
Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
60 GGAAACAAAATTGAACATCTTGCAAAAATTGGATGGAAAAATCATTAACATTCAGTTAT 119  
81 GGAAACAAAATTGAACATCTTGCAAAAATTGGATGGAAAAATCATTAACATTCAGTTAT 140  
120 AACCGTATTCAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179  
141 AACCGTATTCAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 200  
180 GTTTTGAATTTTGAATCTTACATGTTGTCCCACTTCAGATGATGATGATGATGATGATGAT 239  
201 GTTTTGAATTTTGAATCTTACATGTTGTCCCACTTCAGATGATGATGATGATGATGATGAT 260  
240 ATTTGGCAGTGAAGCATTTGTACAGAAATGATGATGATGATGATGATGATGATGATGATGAT 299  
261 ATTTGGCAGTGAAGCATTTGTACAGAAATGATGATGATGATGATGATGATGATGATGATGAT 320  
300 TTGGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359  
321 TTGGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 380  
360 ATGTTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419  
381 ATGTTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 440  
420 ACACCAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 479

Db 441 ACACCAATGATATGACGTAATAGAACTTCACGATTGCTTTCTACCAAGCAACTCCTT 500

QY 480 ACTTATGAAGCACTGGGACTCTGTCTCAGAAAGCAAGGTCACAGCTGGTTGATAGAGA 539

Db 501 ACTTATGAAGCACTGGGACTCTGTCTCAGAAAGCAAGGTCACAGCTGGTTGATAGAGA 560

QY 540 GATTAATACATATGAGAG 556

Db 561 GATTAATACATATGAGAG 577

RESULT 15  
BM552173 1051 bp mRNA linear EST 20-FEB-2002  
LOCUS BM552173  
DEFINITION AGENCOURT\_6543441 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5549545  
5', mRNA sequence.

ACCESSION BM552173  
VERSION BM552173.1 GI:18789814  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1051)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation (LLNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLAM12260 row: e column: 02  
High quality sequence stop: 653.  
Location/Qualifiers  
1..1051  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5549545"  
/issue\_type="leiomysarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. " 6 others

BASE COUNT 318 a 189 c 258 g 280 t

ORIGIN

Query Match 47.3%; Score 470.2; DB 12; Length 1051;  
Best Local Similarity 96.4%; Pred. No. 1.8e-86;  
Matches 502; Conservative 0; Mismatches 15; Indels 4; Gaps 2;

QY 1 CTCACCCAGTGTCTCTCAGATGTTGGGTATGCTGAAAAGACATATGAAAAATATG 60

Db 307 CTCACCCAGTGTCTCTCAGATGTTGGGTATGCTGAAAAGACATATGAAAAATATG 366

QY 61 GAACAAAATTTGAACCTTGAACAAATTTGATGAAAAATCATTAACATTCAGTTAATA 120

Db 367 GAACAAAATTTGAACCTTGAACAAATTTGATGAAAAATCATTAACATTCAGTTAATA 426

QY 121 ACCCGATTCCAGTTCAGATGATATAGATTAGATGAAGTATGATCTTAAAGAG 180

Db 427 ACCCGATTCCAGTTCAGATGATATAGATTAGATGAAGTATGATCTTAAAGAG 486

QY 181 TTTTGTATTTTGTACTATCTTACAAATGTTGCCACTTCAGATGCTGACAGCA 240

Db 487 TTTTGTATTTTGTACTATCTTACAAATGTTGCCACTTCAGATGCTGACAGCA 546

QY 241 TTTGGCCAGTGAAGCATTTGTACAGAAATATAGGCTGCAATCCAAAGCTGTGAAAATT 300

Db 547 TTTGGCCAGTGAAGCATTTGTACAGAAATATAGGCTGCAATCCAAAGCTGTGAAAATT 606

QY 301 TGGCAAGAAAATGATGATGATTTTGGCAAGCTCGTTTGAAGAAAAAGCATTTAATAA 360

Db 607 TGGCAAGAAAATGATGATGATTTTGGCAAGCTCGTTTGAAGAAAAAGCATTTAATAA 666

QY 361 TGGTGGCTTTGATATGAGTAAAGAGCTGCAAGAAAATGCTATGAGAAATCTGGCTGA 420

Db 667 TGGTGGCTTTGATATGAGTAAAGAGCTGCAAGAAAATGCTATGAGAAATCTGGCTGA 726

QY 421 CACCAATGATATGACGTAATAGAACTTCAAGATTGCTTTCTACCAAGCAACT-CCCT 479

Db 727 CACCAATGATATGACGTAATAGAACTTCAAGATTGCTTTCTACCAAGCAACTCCCT 786

QY 480 ACTTATGAAGCACTGGACTCTGT--CCAGAAAGACAAG 517

Db 787 ACTTATGAAGCACTGGACTCTGTCCCAAGAAAGACAAG 827

Search completed: November 27, 2003, 12:28:18  
Job time: 2685.01 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 05:01:24 ; Search time 326.626 Seconds  
(without alignments)  
8215.019 Million cell updates/sec

Title: US-09-835-992A-21

Perfect score: 994  
Sequence: 1 ctaccaccgctgcctccctcag.....ccaccctgggtcgtggtgac 994

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	978	99.4	994	20 AAX40096
2	769	77.4	2663	22 AAH57501
3	769	77.4	3053	23 ABV21847
4	769	77.4	3053	23 ABV27671
5	573.4	57.7	2152	24 AB199762
6	322.8	32.5	432	25 ABX40513
7	266	26.8	1837	23 ABL16479
8	260	26.2	1237	23 ABL16661

9	253	25.5	346	25 ABX42390
10	238	23.9	314	25 ABX35482
11	214	21.5	441529	22 AAX199682
12	211.4	21.3	3613	23 ABL16660
13	211.4	21.3	4082	23 ABL18868
14	211.4	21.3	4082	23 ABL26826
15	208.2	20.9	4356	23 ABL16478
16	199.6	20.1	4003765	22 ABL19683
17	189.8	19.1	412	25 ABX3781
18	106	10.7	330	21 AAC07977
19	99.2	10.0	199	21 AAC07978
20	93.2	9.4	192	15 AAQ7270
21	92	9.3	249	25 ABX26831
22	90.4	9.1	1664976	19 AAV21209
23	71.8	7.2	382	23 AAV12035
24	71.8	7.2	416	23 ABV3180
25	70.6	7.1	349980	22 AAH41226
26	53.2	5.4	705	24 ABT09578
27	52.6	5.3	451	23 ABV42104
28	41.6	4.2	1230	10 AAN90223
29	39.8	4.0	1662	18 AAT75075
30	39.8	4.0	6313	18 AAV74512
31	39	3.9	580073	18 AAT56840
32	38.8	3.9	19387	22 AAS34561
33	38.6	3.9	9207	22 AAS46789
34	38.6	3.9	9207	24 ABK28433
35	38.4	3.9	327	23 ABV05120
36	38.4	3.9	2557	23 ABL15768
37	38.4	3.9	3965	23 ABL02948
38	38.2	3.8	5189	23 ABL25648
39	38.2	3.8	1659	23 AAS51860
40	38.2	3.8	1662	23 AAS54694
41	38	3.8	10279	24 ABL29276
42	38	3.8	10279	24 ABL35590
43	37.8	3.8	10279	24 AAD2327
44	37.8	3.8	1353	21 AAA51780
45	37.8	3.8	1664976	19 AAV21209

#### ALIGNMENTS

RESULT 1	
AAX40096	
ID AAX40096 standard; DNA; 994 BP.	
XX	
AC AAX40096;	
XX	
DT 02-JUL-1999 (first entry)	
XX	
DE Gastric cancer associated gene.	
XX	
KW Cancer associated antigen; diagnosis; research; treatment; human;	
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KW prostate cancer; ss.	
XX	
OS Homo sapiens.	
XX	
PN WO9904265-A2.	
XX	
PD 28-JAN-1999.	
XX	
PF 15-JUL-1998;	98WO-US14679.
XX	
PR 22-JUN-1998;	98US-0102322.
PR 17-JUL-1997;	97US-0896164.
PR 10-OCT-1997;	97US-0061599.
PR 10-OCT-1997;	97US-0061765.
PR 10-OCT-1997;	97US-0948705.
PR 11-OCT-1997;	97GB-0021697.
XX	
PA (LUDW-) LUDWIG INST CANCER RES.	
XX	

Bovine EST associa  
Bovine EST associa  
Mycobacterium tube  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Mycobacterium tube  
Bovine EST associa  
Human secreted pro  
Human secreted pro  
Human genome fragm  
Human GDP-mannose  
Methanococcus jan  
Human prostate exp  
Human prostate exp  
Pyrococcus abyssi  
Phase-1 Rat CT gen  
Human prostate exp  
Malaria-specific D  
DNA encoding argin  
Staphylococcus aur  
Mycoplasma genital  
Human DNA for a no  
Tumour suppressor  
DNA transcription  
Human prostate exp  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Staphylococcus aur  
Staphylococcus aur  
Chemically treated  
Human immune syste  
Chemically treated  
S. cerevisiae esse  
Methanococcus jan



CC pharmaceutical agents. Expression of (I) in a sample indicates the  
 CC differentiation of embryonic stem cells into a tissue selected from  
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
 CC tissues. (I) and (II) are used to produce an expression profile that  
 CC defines a metabolic or developmental process, treatment, condition,  
 CC disease or disorder. The gene profile can be used for diagnosis,  
 CC prognosis or monitoring of treatments and for investigating a  
 CC predisposition to a disorder where the gene is associated with a  
 CC cancer, immunopathology or neuropathology.

XX Sequence 2663 BP; 807 A; 472 C; 599 G; 785 T; 0 other;

Query Match 77.4%; Score 769; DB 22; Length 2663;  
 Best Local Similarity 91.3%; Pred. No. 9.1e-207;  
 Matches 899; Conservative 0; Mismatches 66; Indels 20; Gaps 8;

QY 1 CTCACCCAGTGGCTCTCAGATGTTGGGTATGCTGGAAGAAAGACATATGAAAAATATG 60  
 DB 566 CTCACCCAGTGGCTCTCAGATGTTGGGTATGCTGGAAGAAAGACATATGAAAAATATG 625  
 QY 61 GAACAAAAATTAACACTTTCGAAAAATTCGAAAAATTCATTAACCTTCAGTTAATA 120  
 DB 626 GAACAAAAATTAACACTTTCGAAAAATTCGAAAAATTCATTAACCTTCAGTTAATA 685  
 QY 121 ACCCGTATTCAGTTCGAAAGATGAATACAGTTTATGATGATGATGATGATGATGATGATG 180  
 DB 686 ACCCGTATTCAGTTCGAAAGATGAATACAGTTTATGATGATGATGATGATGATGATGATG 745  
 QY 181 TTTTGTATTTTGAATCTATCTTACAAATGTTGCCAATTCAGATGGTGTGAGCAGCA 240  
 DB 746 TTTTGTATTTTGAATCTATCTTACAAATGTTGCCAATTCAGATGGTGTGAGCAGCA 805  
 QY 241 TTTTGGCCAGTGAACATTTTGAACAGATATGCGCTGCAATCCAAAGCTGTGGAATTT 300  
 DB 806 TTTTGGCCAGTGAACATTTTGAACAGATATGCGCTGCAATCCAAAGCTGTGGAATTT 865  
 QY 301 TGGCAACAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
 DB 866 TGGCAACAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 925  
 QY 361 TGGTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
 DB 926 TGGTGGCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 985  
 QY 421 CACCAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 DB 986 CACCAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1045  
 QY 481 CTTATGAAGCATGCGGACTGTGCCAAGAGCAAGGTGCAAGCGCTGGTGTATGAGAGAG 540  
 DB 1046 CTTATGAAGCATGCGGACTGTGCCAAGAGCAAGGTGCAAGCGCTGGTGTATGAGAGAG 1105  
 QY 541 ATATATCATATGAGAGAAAGTGGTCTATTAATCTATGATGATGATGATGATGATGATGATG 600  
 DB 1106 ATATATCATATGAGAGAAAGTGGTCTATTAATCTATGATGATGATGATGATGATGATGATG 1165  
 QY 601 ACCCACTAGGCGCTACAGGTCTTTGCTCAAGTGCAGAACTTGTCTGCGCACTGAGAGAGG 660  
 DB 1166 ACCCACTAGGCGCTACAGGTCTTTGCTCAAGTGCAGAACTTGTCTGCGCACTGAGAGAGG 1225  
 QY 661 AAGCGGAAAGAGAGCAAGGTCTTGTGTCAGAAAGGTGCTGCGCACTGATTTTATTTATGAT 720  
 DB 1226 AAGCGGAAAGAGAGCAAGGTCTTGTGTCAGAAAGGTGCTGCGCACTGATTTTATTTATGAT 1283  
 QY 721 TGGAGAACTGTGTGTTGAACACTTACAAATGAGGTTTTCGCGAGCCGCACTGATTC 780  
 DB 1284 TGGAGAGAGCTGTGTGTTGAACACTTACAAATGAGG--TTTTCGCGAGCCGCACTGATTC 1340  
 QY 781 TTTTGAAGTATCAAAATTTGAAGCGTTCACCAAGCTCTGCAAGTATTAATTTTAA 840  
 DB 1341 TTTTGAAGTATCAAAATTTGAAGCGTTCACCAAGCTCTGCAAGTATTAATTTTAA 1397  
 QY 841 NNNAAATCTGTTTAAAGGAGATGAGAAAGAACTTAAAGGAGGAGGAGACA 900

DB 1398 GCGAAATCTGTTTAA--GGAATGAGAAAGAACTTGA---AGAGAAAGGAGACA 1451  
 QY 901 ATTGTGAAGAAATTCGAGGAAATTTTCCCTTCAGAGGAAANAATGCCCCCTGG 960  
 DB 1452 GTTGTGAAGAAATTCGAGGAAATTTTCCCTTCAGAGGAAANAATGCCCCCTGG 1505  
 QY 961 GGGGTAAAGAGAGCCACCTGGGG 985  
 DB 1506 GGGTAAAGAGAGCCACCTGGGTG 1530

RESULT 3  
 ABV21847  
 ID ABV21847 standard; cDNA; 3053 BP.

ABV21847;  
 13-SEP-2002 (first entry)

Human prostate expression marker cDNA 21838.

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 pharmacogenomic marker; gene; ss.

Homo sapiens.

WO200160860-A2.

23-AUG-2001.

20-FEB-2001; 2001WO-US05171.

17-FEB-2000; 2000US-18319P.

16-MAR-2000; 2000US-189862P.

25-MAY-2000; 2000US-207454P.

09-JUN-2000; 2000US-211314P.

18-JUL-2000; 2000US-219007P.

13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE;

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of

prostate cells and correlating with presence of prostate cancer, useful

for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 3703; 11750bp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising

a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

specification or its complement. (I) is useful for:

(a) assessing whether a patient is afflicted with prostate cancer;

(b) monitoring the progression of prostate cancer in a patient;

(c) assessing the efficacy of a test compound to inhibit prostate

cancer in a patient;

(d) assessing the efficacy of a therapy for inhibiting prostate cancer

in a patient;

(e) selecting a composition for inhibiting prostate cancer in a patient;

(f) assessing the prostate cell carcinogenic potential of a compound;

(g) determining whether prostate cancer has metastasized in a patient;

(h) assessing the aggressiveness or indolence of prostate cancer in a

patient;

(I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 3053 BP; 1010 A; 572 C; 887 G; 580 T; 4 other;

Query Match 77.4%; Score 769; DB 23; Length 3053;  
 Best Local Similarity 91.3%; Pred. No. 9.6e-207;  
 Matches 899; Conservative 0; Mismatches 66; Indels 20; Gaps 8;

```

QY 1 CTCACCCAGTGTGCTCTCAGATGTTGGGTATCTGAAAAAGAACATATGAAAAATATG 60
DB 588 CTCACCCAGTGTGCTCTCAGATGTTGGGTATGCTGAAAAAGAACATATGAAAAATATG 647
QY 61 GAACAAAAATTTGAACACTTTTGCAAAAAATGATGAAAAATCATAAACATTCAGTTAATA 120
DB 648 GAACAAAAATTTGAACACTTTTGCAAAAAATGATGAAAAATCATAAACATTCAGTTAATA 707
QY 121 ACCCGTATTTCCAGTTCAGAGATGAATACAGTTTAGATGAAGTATGAGCATCTAAAGAG 180
DB 708 ACCCGTATTTCCAGTTCAGAGATGAATACAGTTTAGATGAAGTATGAGCATCTAAAGAG 767
QY 181 TTTTGGATTTTGGATCTTACATATGTTGCTCCACTTCAGATGCTGTCAGACGCA 240
DB 768 TTTTGGATTTTGGATCTTACATATGTTGCTCCACTTCAGATGCTGTCAGACGCA 827
QY 241 TTTTGGCAGTGAAGATTTGTACAGAGATAGGCTGCAATCCAAAGCTGTGAAATTT 300
DB 828 TTTTGGCAGTGAAGATTTGTACAGAGATAGGCTGCAATCCAAAGCTGTGAAATTT 887
QY 301 TGGCACAAGAAATGATGACTGATTTTGCAGAGCTCGTTGAGAAAAAGCATTTATTAAAA 360
DB 888 TGGCACAAGAAATGATGACTGATTTTGCAGAGCTCGTTGAGAAAAAGCATTTATTAAAA 947
QY 361 TGGTGGCTTTGATATGATGATTAAGAAAGCTGCAAGAAAAATGCTATGGAATCTGCGCTGA 420
DB 948 TGGTGGCTTTGATATGATGATTAAGAAAGCTGCAAGAAAAATGCTATGGAATCTGCGCTGA 1007
QY 421 CACCAATATATTTAGATGATTAAGAACTTCAGATTTCTTCTACCAAGAACTCTCTTA 480
DB 1008 CACCAATATATTTAGATGATTAAGAACTTCAGATTTCTTCTACCAAGAACTCTCTTA 1067
QY 481 CTTATGAACACTGGGACTCTGTCAGAAAGAACAGTGCACACGCTGTTGATAGAGAG 540
DB 1068 CTTATGAACACTGGGACTCTGTCAGAAAGAACAGTGCACACGCTGTTGATAGAGAG 1127
QY 541 ATATATCATATGAGAGAAAGTGGGTATTAATCTCTAGTGTGATGATTTCAAAGGAC 600
DB 1128 ATATATCATATGAGAGAAAGTGGGTATTAATCTCTAGTGTGATGATTTCAAAGGAC 1187
QY 601 ACCCACTAGGCGCTACAGGCTTGTGCTCAGTGTGACAGAACTCTGCTGGCAGCTGAGAGGG 660
DB 1188 ACCCACTAGGCGCTACAGGCTTGTGCTCAGTGTGACAGAACTCTGCTGGCAGCTGAGAGGG 1247
QY 661 AAGCCGGAAGAGGCAAAAGTCTGCTGCAAAAGTGTGCTCTGCGCATTAATTTANGAT 720
DB 1248 AAGCCGGAAGAGGCAAAAGTCTGCTGCAAAAGTGTGCTCTGCGCATTAATTTANGAT 1305
QY 721 TGGAGGAACCTGTGTGTTGAACACTCTACAAAGATGGGTTTCCCGAAGCCCGCACTTCC 780
DB 1306 TGGAGGAACCTGTGTGTTGAACACTCTACAAAGATGGGTTTCCCGAAGCCCGCACTTCC 1362
QY 781 TTTTGAACCTCATCAAAATTTGAAGCNGTTCACCAAGCTCTGCAAGTATNGGTTTAA 840
DB 1363 TTTTGAACCTCATCAAAATTTGAAGCNGTTCACCAAGCTCTGCAAGTATNGGTTTAA 1419
QY 841 NGNAAATCTNGTTTTTAAAGNGATTTGAAGAGAAACNTNAAAGAGGAGANGGGAGACA 900
DB 1420 GCGCAATCTGTTTTTAA--GGAGATTTGAAGAGAAACNTNAAAGAGGAGANGGGAGACA 1473
QY 901 ATTGTGGAAGAAAAATNGGNGGGAATTTTCCCTTCAAGGGAAAAAATGCGCCCTGG 960
DB 1474 GTTTTGAAGAAAAATGCGTGTGTTT--ATTTTGTGCTTCAGGTGAAG--ATGCGCCCTGG 1527
QY 961 GGGGTAAAGANGGCCACCTGGGG 985
DB 1528 GGGTAAAGANGGCCACCTGGGTGTG 1552

```

RESULT 4  
ABV27671  
ID ABV27671, standard; cDNA; 3053 BP.

```

XX ABV27671;
AC 16-SEP-2002 (first entry)
XX
DT Human prostate expression marker CDNA 27662.
XX
DE Human; prostate cancer; cyrostatic; carcinogen; pharmacodynamic marker;
XX phamacogenomic marker; gene; ss.
XX Homo sapiens.
XX
OS WO200160860-A2.
XX
PN 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
XX
PR 16-MAR-2000; 2000US-189862P.
XX
PR 25-MAY-2000; 2000US-207454P.
XX
PR 09-JUN-2000; 2000US-211314P.
XX
PR 18-JUL-2000; 2000US-219007P.
XX
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR MPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
PS Claim 1; Page 5667; 11750p; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 3053 BP; 1010 A; 572 C; 887 G; 580 T; 4 other;
XX
Query Match 77.4%; Score 769; DB 23; Length 3053;
Best Local Similarity 91.3%; Fred. No. 9.6e-207;
Matches 899; Conservative 0; Mismatches 66; Indels 20; Gaps 8;
QY 1 CTCACCCAGTGTGCTCTCAGATGTTGGGTATGCTGAAAAAGAACATATGAAAAATATG 60
DB 588 CTCACCCAGTGTGCTCTCAGATGTTGGGTATGCTGAAAAAGAACATATGAAAAATATG 647
QY 61 GAACAAAAATTTGAACACTTTTGCAAAAAATGATGAAAAATCATAAACATTCAGTTAATA 120
DB 648 GAACAAAAATTTGAACACTTTTGCAAAAAATGATGAAAAATCATAAACATTCAGTTAATA 707
QY 121 ACCCGTATTTCCAGTTCAGAGATGAATACAGTTTAGATGAAGTATGAGCATCTAAAGAG 180
DB 708 ACCCGTATTTCCAGTTCAGAGATGAATACAGTTTAGATGAAGTATGAGCATCTAAAGAG 767
QY 181 TTTTGGATTTTGGATCTTACATATGTTGCTCCACTTCAGATGCTGTCAGACGCA 240

```

Db 768 TTTTGGATTTTGGACTATCTTACATGTTGTCCACCTTACAGATGCTGCAGACAGCA 827  
 Oy 241 TTTTGGCCAGTACACATTTTGTACAGAAATATGCGCTTGCATTCGAAAGTGTGAAATTT 300  
 Db 828 TTTTGGCCAGTACACATTTTGTACAGAAATATGCGCTTGCATTCGAAAGTGTGAAATTT 887  
 Oy 301 TGGCACAAGAAATGATGACTGATTTGGCCAAAGCTCGTTTGAAGAAAAACATTTATTA 360  
 Db 888 TGGCACAAGAAATGATGACTGATTTGGCCAAAGCTCGTTTGAAGAAAAACATTTATTA 947  
 Oy 361 TGGTGGCTTGTATATGATTAAGAAAGCTGCAAGAAATGCTATAGAAATCTGCGCTGA 420  
 Db 948 TGGTGGCTTGTATATGATTAAGAAAGCTGCAAGAAATGCTATAGAAATCTGCGCTGA 1007  
 Oy 421 CACCAATGATATTGACGTAATAGAACTTACGATTTGCTTTCTACCAAGAACTCTTGA 480  
 Db 1008 CACCAATGATATTGACGTAATAGAACTTACGATTTGCTTTCTACCAAGAACTCTTGA 1067  
 Oy 481 CTTATGAAGCACTGGGACTCTGTCCAGAAAGACAAAGGTGCAACGGCTGGTGTATAGAGAG 540  
 Db 1068 CTTATGAAGCACTGGGACTCTGTCCAGAAAGACAAAGGTGCAACGGCTGGTGTATAGAGAG 1127  
 Oy 541 ATATATCATATGAGAGAAAGTGGGTCTATTAATCTATGCTGAGTGAATTTCAAGAGAC 600  
 Db 1128 ATATATCATATGAGAGAAAGTGGGTCTATTAATCTATGCTGAGTGAATTTCAAGAGAC 1187  
 Oy 601 ACCCACTAGGCGCTTACAGGTCTTGTCTCACTGTCAGAACTCTGCTGGCAGCTGAGAGGGG 660  
 Db 1188 ACCCACTAGGCGCTTACAGGTCTTGTCTCACTGTCAGAACTCTGCTGGCAGCTGAGAGGGG 1247  
 Oy 661 AAGCCGGAAGAAAGCAAGGCTCTGTGTCAGAAAGGCGTCTGCGCATTAATTTANGCAT 720  
 Db 1248 AAGCCGGAAGAAAGCAAGGCTCTGTGTCAGAAAGGCGTCTGCGCATTAATTTANGCAT 1305  
 Oy 721 TGGAGCACTGTGTGTGTATCACTCTACAGATGAGGGTGTTCGCCGAAGCCGCCAGTTCC 780  
 Db 1306 TGGAGCACTGTGTGTGTATCACTCTACAGATGAGGGTGTTCGCCGAAGCCGCCAGTTCC 1362  
 Oy 781 TTTTGAACCTCATCAAAATTGAAGCCNGTTCACACCAAGCTTGCAGATGATNGTTTAA 840  
 Db 1363 TTTTGAACCTCATCAAAATTGAAGCCNGTTCACACCAAGCTTGCAGATGATNGTTTAA 1419  
 Oy 841 NGNAATATCTNGTTTAAAGNGGATGAGAGAAACNTNAAGGGGAAGGGGAACA 900  
 Db 1420 GGCATAATCTGTTTAA--GGAGATGAGAAACTTGA---AGAGAAAGGGGAACA 1473  
 Oy 901 ATTGTGAAGAAATATCGNGGGAATTTTGCCTTCAAGGGAAATATGCGCTGG 960  
 Db 1474 GTTGTGAAGAAATATCGNGGT----ATTTTGCCTTCAAGGTGAAG-ATGGCCCTGG 1527  
 Oy 961 GGGGTAAAGANGCCACCTGGGG 985  
 Db 1528 GGGTAAAGAGGCCACCTGGGTGTG 1552

RESULT 5  
 ABI99762  
 ID ABI99762 standard; cDNA; 2152 BP.  
 AC ABI99762;  
 XX

DT 07-MAR-2002 (first entry)  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:842.  
 XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
 OS Mus musculus.  
 PN WO200188188-A2.  
 XX 22-NOV-2001.

XX 18-MAY-2001; 2001MO-JP04192.  
 PF 18-MAY-2000; 2000JP-0145977.  
 PR (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
 PA Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
 DR WPI, 2002-034733/04.  
 DR P-PSDB; AB857301.  
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or  
 PT by determining the expression profile of a gene group comprising these  
 PT genes.  
 PS Claim 2; Page 2087-2091; 2690pp; English.  
 CC The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (I) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (I). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (ABI99202 to ABI99912, encoding  
 CC the protein sequences in AB857020 to AB857374) or by determining the  
 CC expression profile of a gene group comprising these genes. The  
 CC expression levels or expression profiles produced by these genes are  
 CC used as an indicator when screening for ischaemic condition-improving  
 CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914  
 CC represent PCR primers for a mouse ischaemic condition related sequence,  
 CC which are used in the exemplification of the present invention.  
 XX  
 XX Sequence 2152 BP; 586 A; 455 C; 576 G; 535 T; 0 other;  
 Query Match 57.7%; Score 573.4; DB 24; Length 2152;  
 Beef Local Similarity 78.9%; Pred. No. 1.6e-151;  
 Matches 765; Conservative 0; Mismatches 186; Indels 19; Gaps 7;  
 Oy 1 CTCACCGAGTTGCTCTCAGATGTTGGGTATGCTGTAAGAAACATATGAAAAATATG 60  
 Db 479 CACATCCGATTAATCTCTCAGATGTTGGGTATGCTGTAAGAAACATATGAAAAATATG 538  
 Oy 61 GAACAAAATTTGAACACTTTTGCAAAAATTTGATGAGAAATATATTAACATTCAATTATA 120  
 Db 539 GAACAAAATTTGAACACTTTTGCAAAAATTTGATGAGAAATATATTAACATTCAATTATA 598  
 Oy 121 ACCGATATCCAGTTCAGATGATGATACAGTTTATGATGAAAGTATGCGATTAAGAAG 180  
 Db 599 ACAGTATATCCAGTTCAGATGATGATACAGTTTATGATGAAAGTATGCGATTAAGAAG 658  
 Oy 181 TTTTGAATTTTGTGACTATCTTACATATGTTGCTCCACTTCAGATGCTGCGACAGCA 240  
 Db 659 TTTTGAATTTTGTGACTATCTTACATATGTTGCTCCACTTCAGATGCTGCGCGCGCA 718  
 Oy 241 TTTTGGCCAGTGAAGCAATTTGTACAGAGATATGCGCTGCATCAAGCTGTGAAATTT 300  
 Db 719 TTTTGGCCAGTGAAGCAATTTGTACAGAGATATGCGCTGCATCAAGCTGTGAAATTT 360  
 Oy 301 TGGCACAAGAAATGATGACTGATTTGGCCAAAGCTCGTTTGAAGAAAAACATTTATTA 420  
 Db 779 TGGCCACAAGAAATGATGACTGATTTGGCCAAAGCTCGTTTGAAGAAAAACATTTATTA 480  
 Oy 361 TGGTGGCTTGTATATGATTAAGAAAGCTGCAAGAAATGCTATAGAAATCTGCGCTGA 480  
 Db 839 TGGTGGCTTGTATATGATTAAGAAAGCTGCAAGAAATGCTATAGAAATCTGCGCTGA 898  
 Oy 421 CACCAATGATATTGACGTAATAGAACTTACGATTTGCTTTCTACCAAGAACTCTTGA 940  
 Db 899 CACCAATGATATTGACGTAATAGAACTTACGATTTGCTTTCTACCAAGAACTCTTGA 958  
 Oy 481 CTTATGAAGCACTGGGACTCTGTCCAGAAAGACAAAGGTGCAACGGCTGGTGTATAGAGAG 540

```

DB 959 CTTAGACACACTGGGGCTCTGCTCCAGAAAGACAGTGGAACTCTGGTGGACAGAGGGG 1018
QY 541 ATATATCATATGAGAGAAAGTGGTGCATATATCTAGTGGTGCATGATTTTCAAGGGAC 600
DB 1019 ACACACTTACGAGAGAAAGTGGTGCATCAACCTTAGTGGAGGCTCATCTTCAAGGGAC 1078
QY 601 ACCCACTAGCCCTAAGTCTTGTCTCATGTGGCAAACTCTGTCTGGCAAGCTGAGGGG 660
DB 1079 ACCCACTGGGCGCAACAGGTCTGGCTCAGTGGCGGAGCTCTGTGGCAAGCTGAGGGG 1138
QY 661 AAGCCGGAAGAGAGCAAGTCTCTGTGCAGAAAGTGGTCTGTGCAGCATATTTANGCAT 720
DB 1139 AAGCCGGAAG--GAGGAGAGTTCGCGGGCAAGAGTGGCTCTGACAGCAATCTAGGCTC 1196
QY 721 TGGAGAACTGTGTGTGTAACACTTCAAGATGGGGTTTCCCGAGACCGCACTTCC 780
DB 1197 CGAGGAGAGTGTGTGTGTAACCTTCAAGATGGG--TTTTCGCGAAGCTGCACTCC 1254
QY 781 TTTTGAATCATCAAAATTTGAAGCCNGTTCACCAACTCTGTGCAGATNATNGTTTAA 840
DB 1255 TTTCAGAG--CGACACAGGTTTCAAGCTGCCA--CCAGCTGTGAGGGGATGATTTAA 1310
QY 841 NGNAAATCTNGTTTAAAGNGATTGAGAGAAACNTNAAAGAGAGAGAGGAGACA 900
DB 1311 GGCACCACTGCTCTTAA--GGAGATTGAGAGAGAGCTTGA---AGAGAGAGGAGACA 1364
QY 901 ATTTGTGAAGAAATTCGNGGGGAATTTTGCCTTCAAGGGGAAANAATGGCCCTGG 960
DB 1365 GTTCGGAAGAAATTCGTGGC-----ATTTTGTCTTCAAGTGAAGATGGCCCTGGA 1419
QY 961 GGGGTAAAG 970
DB 1420 GGCAGAGAG 1429

RESULT 6
ABX40513
ID ABX40513 standard; cDNA, 432 BP.
XX
AC ABX40513;
XX
DT 20-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #5678.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US200237139-A1.
XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-0960352.
XX
PR 12-JAN-1999; 99US-115707P.
XX
PR 11-JAN-2000; 2000US-0480902.
XX
PA (BYAT/) BYATT J C.
PA (WATH/) WATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Machialagan N, Tao N, Warren WC;
PI WPI; 2003-110599/10.
XX
PT New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle

```

```

XX
PS Claim 2; SEQ ID No 5678; 245bp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMPD), derived
CC from cattle, and the LMPD nucleic acid can specifically hybridize to a
CC second nucleic acid molecule comprising any of 1512 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements thereof.
CC Also included are: (1) a transformed cell having a nucleic acid
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
CC translating sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 1512 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridization between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMPD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 1512 bovine
CC LMPD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docid=20020137139.
XX
SQ Sequence 432 BP; 127 A; 75 C; 122 G; 108 T; 0 other;
XX
Query Match 32.5%; Score 322.8; DB 25; Length 432;
Best Local Similarity 88.7%; Pred. No. 5,1e-81;
Matches 360; Conservative 0; Mismatches 44; Indels 2; Gaps 1;

QY 357 AAATGCTGTGCTTGTGATATGATGTAAGAGTGAAGAAATGCTATGAGAAATCTGGC 416
DB 2 AAATGCTGTGCTTGTGATATGATGTAAGAGTGAAGAGTGAAGAGTGAAGAAATCTGGC 61
QY 417 CTGACACCAATATGATATGACGTATATAGAACTTACAGATTGCTTTTCTACCAAGCACTC 476
DB 62 CTGACACCAAGATATGATATGACGTATATAGAACTTACAGATTGCTTTTCTGCAATGACTT 121
QY 477 CTACTTATGAGCACTGGGACTCTGTCGAGAGGACAAAGTGAACCGTGTATATGA 536
DB 122 ATTTACTTACGAAGCACTGGGACTGTGTCAGAGGTCAGAGGTGAAGAACTGCTGAAAGA 181
QY 537 GGAATATATACATATGAGAGAAAGTGGTCAATTAATCTTAGTGGAGCTGATTTCAAG 596
DB 182 GGAGATATATCTTATGAGAGAAAGTGGTCAATTAATCTTAGTGGAGCTGATTTCAAG 241
QY 597 GGACACCACTAGGCGCTTACAGGTCTTCTCAGTGTGAGAACTGTGTGAGCTGAGAG 656
DB 242 GGGCACCACTTGTGAGCTTACAGGTCTGCTCAGTGTGAGAACTGTGTGAGCTGAGAG 301
QY 657 GGGGAAGCCGGAAGAGGCAAGTCTGTGTGCAAGAGTGTGTGCAAGCTGCAATTTAN 716
DB 302 GGGGAAGCCGGAAG--AAGGCAAGTCTGTGTGCAAGAGTGTGTGCAAGCTGCAATTTAN 359
QY 717 GCATTGAGAGAACTGTGTGTAACACTTCAAAATGGGCTTTC 762
DB 360 GCATTGAGAGAACTGTGTGTGTAACACTTCAAAATGGGCTTTC 405

RESULT 7
ABL16479
ID ABL16479 standard; DNA, 1837 BP.
XX
AC ABL16479;
XX
DT 26-MAR-2002 (first entry)

```

XX	DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 910.
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;
XX	KV	pharmaceutical; gene; ds.
XX	OS	Drosophila melanogaster.
XX	PN	WO200171042-A2.
XX	PD	27-SEP-2001.
XX	PF	23-MAR-2001; 2001WO-US09231.
XX	PR	23-MAR-2000; 2000US-191637P.
XX	PR	11-JUL-2000; 2000US-0614150.
XX	PA	(PEKE ) PE CORP NY.
XX	XX	
XX	PI	Venter JC, Adams M, Li PMD, Myers EW;
XX	DR	WPI; 2001-656860/75.
XX	XX	
PT	XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	XX	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	XX	interactions -
PS	XX	
PS	XX	Claim 1; SEQ ID NO 910; 21pp + Sequence Listing; English.
XX	CC	
CC	CC	The invention relates to an isolated nucleic acid detection reagent
CC	CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	CC	useful in developmental biology and in elucidating cell signalling and
CC	CC	cell-cell interactions in higher eukaryotes for the development of
CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	CC	discloses genomic DNA sequences (AB16176-AB16177), expressed DNA
CC	CC	sequences (AB161840-AB161841) and the encoded proteins
CC	CC	(AB1617737-AB1617738).
CC	CC	The sequence data for this patent did not form part of the printed
CC	CC	specification, but was obtained in electronic format directly from WIPO
CC	CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	XX	
XX	XX	Sequence 1837 BP; 428 A; 464 C; 528 G; 417 T; 0 other;
XX	XX	
XX	XX	Query Match 26.8%; Score 266; DB 23; Length 1837;
XX	XX	Best Local Similarity 61.0%; Pred. No. 1.1e-64;
XX	XX	Matches 465; Conservative 0; Mismatches 292; Indels 5; Gaps 2
OY	1	CTACCCAGTTGCTCTCTCAGATGTTGGGTATCTGGAAAAGACATATGAAAAATATG 60
DB	553	CTGGCCCAATGCTGCCAGATCTTCGGCAATGCCGCAAGGACACATGAAGAATATG 612
OY	61	GACCAAAATGGAACCTTTGGAATAATGATGAAAAAATATATAACCTTCAGTTAAT 120
DB	613	GCATTAAGCCCGAGCATTTTCGGCAAGATCGCTTGAAGAACACAAAGCATCTCTCAATA 672
OY	121	ACCGTATTCAGTTCACAGATGATACAGTTTATGATGAATGATGATTAAGAAG 180
DB	673	ATCCTTACTCGAAGTTCCGCAATGATACACCTTGAAGAGATTTATGAAGTGGCCACAG 732
OY	181	TTTTTGA---TTTTTGACTATCTTACAAATGTTGTCCTCACTTCAGATGCTGTCGACAG 237
DB	733	TTGTGGAAGGAGTGTGACCAAGCTGACAGTGTGTTGCCACTTCCGATGATCTGGAGCTG 792
OY	238	CAATTTTGGCCAGTGAAGCATTTTGTACAGAAAGTATAGGCTGCAATCCAAAGCTGTGGA 297
DB	793	CCATTTCTGCTTCGAGGCTTTGTCGTCGTGCGCACAGGATTTGAAAAGCGCTGTGCGAA 852
OY	298	TTTGTGCAACAAGAAATGATGATGATTTTGGCAAGCTCGTTTGAAGAAAAAAGCATTAAT 357
DB	853	TTGTAGGATGAGAGATGCGCAATGACCGGCGCTCCACCTTTGCGCAAGAGCTTGATGA 912
OY	358	AAATGTTGGCTTTTATATGATTAAGAAAGCTGCAAGAAATGCTATGAGAAATCTGGCC 417

Dd	913	AGATTGCTGGAAACCGAATATGACCCGCTGGCCACCACCCAGCTCTTTCGCAAAAGTGGCT	972
Qy	418	TGACACCAAAATGATATTGACGTAATAGAACTTCACGATTCGTTTCTACCAACGAACTCC	477
Dd	973	ACAAGCCCAAGGATGTCCAGGTTGTGGAATCGACAGATTGCTTTCGGCCAAAGATTGA	1032
Qy	478	TTACTTATGAAGCACTGGGACTCTGTCCAGAGACAAAGTGCACACCTGTTATATGAG	537
Dd	1033	TCAGGTATGAGGCACTTGGACTGTGTGAGAGGGCCAAACGCCGGAGATTCAATCAGCCTG	1092
Qy	538	GAGATTAATACATATGAGGAAAGTGGTCAATTAATCCTAGTGGTGAATTTCAAAAG	597
Dd	1093	GAGACAACACTTACGGTGGCAAGTTTCGTGGTCAACCCCAAGTGGTGTGATTCGAAAG	1152
Qy	598	GACACCCACTAGGCGCTACAGGTTCTCTCAGTGTGAGAACTCTGCTGGCAGCTGAGAG	657
Dd	1153	GCCATCCTCTGTGGGGGCCACGGGTGTGGACAATGTGCTGAGCTCTGCTGCAGCTCCGTG	1212
Qy	658	GGGAGCCCGGAAAAAGGCAAAAGTTCTCTGTGCAAGGTGGCTCTGCNGCATTAATTANG	717
Dd	1213	GATTGGCTG--AGAAGCGCCAGGTGCCCAATGCTCAGTTGGCTTTGCAGACAAATCGGG	1270
Qy	718	CATTGAGGAAGACTGTGTTGTAACACTTCACAGATGGCGTT	759
Dd	1271	ACTGGGTGTGTGCCGTGTGTGTGCTTGTATCGTCTGGGTTT	1312
RESULT 8			
ABL16661	ABL16661 standard; DNA, 1237 BP.		
ID	ABL16661		
XX	AC		
XX	ABL16661;		
XX	26-MAR-2002 (first entry)		
XX	Drosophila melanogaster genomic polynucleotide SEQ ID NO 1456.		
XX	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ds.		
XX	Drosophila melanogaster.		
OS	Drosophila melanogaster.		
XX	WO200171042-A2.		
XX	27-SEP-2001.		
PD	23-MAR-2001; 2001WO-US09231.		
PF	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX	(PEKE ) PE CORP NY.		
PA	Ventec JC, Adams M, Li PWD, Myers EW;		
PI	WPI; 2001-656860/75.		
DR	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions -		
XX	Claim 1; SEQ ID NO 1456; 21pp + Sequence Listing; English.		
PS	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
CC	sequences (ABL01840-ABL16175) and the encoded proteins		
CC	(ABB57737-ABB72072).		
CC	The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIFO		

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1237 BP; 289 A; 315 C; 369 G; 264 T; 0 other;

Query Match 26.2%; Score 260; DB 23; Length 1237;  
Best Local Similarity 60.7%; Pred. No. 4.6e-63;  
Matches 459; Conservative 0; Mismatches 292; Indels 5; Gaps 2;

```

QY 7 CAGTCTCTCTCAGATGTTGGTATGCTGAGAAAAGAACATATGAAACAA 66
DB 394 CATTTGCCCAAGATCTTCGCGACGCGGCAAGAGACATGAAGATGAGCACTA 453
QY 67 AAATTGAACATTTGCAAAATATGATGAGAAAATATTAACCTAATACCGGT 126
DB 454 AGCCGAGCATTTGCGCAAGATGCTTGGAGAACACAGCACTCCGTCATATACCT 513
QY 127 ATTCCAGTTCCAGATGATACAGTTTATGATGATGATGATGATGATGATG 186
DB 514 ACTCGCAGTTCCGCGATGATACACCTGAGCAGATTTAGATGATGATGATG 573
QY 187 A---TTTTTGAATATCTTACATATGTTTCCACTTCAGATGCTGACAGCAATTT 243
DB 574 AGGAGTGTGACCAAGCTGAGTGTGCTGCTCCACTTCGATGATGATGATG 633
QY 244 TGGCCAGTGAAGCATTTGTACAGAACTATGAGCTGCAATCCAAAGCTGGAATTTGG 303
DB 634 TGCCCTCCAGGCTTCGTGCGTCCGACGAGATGAGAAAGCGCTGTGGAGATTGG 693
QY 304 CACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 363
DB 694 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 753
QY 364 TTGGCTTATATGATGATGATGATGATGATGATGATGATGATGATGATG 423
DB 754 CTGGAAACCATATGACCCGCTGCGCCACCGAGCTGTGTCGCAAGTGGGTACAGC 813
QY 424 CAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATG 483
DB 814 CCGAGATGCTCAGTGTGATGATGATGATGATGATGATGATGATGATGATG 873
QY 484 ATGAAACATGAGATCTCTGTCAGAGAGACAAAGTCAACGCTGATGATGATG 543
DB 874 ATGAGCATTGAGATCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 933
QY 544 ATACATATGAGAGAAAGTGGTCAATATCTTATGCTGATGATGATGATGATG 603
DB 934 ACACCTACCGTGTGAGATCTGTCGATGATGATGATGATGATGATGATGATG 993
QY 604 CACTAGGCGCTACAGTCTTGTGCTGATGATGATGATGATGATGATGATGATG 663
DB 994 CTCTGGGCGCAGGCTGTGAGCAATGATGATGATGATGATGATGATGATGATG 1053
QY 664 CCGGAAAGAGAGCAAAATGCTGTCGTCAGAGTGTGCTGTCGTCAGATTTT 723
DB 1054 CTG--AGAGGCGCAGATGCTCAATGCTCAGTTGGTCTGTCGAGCAATCTG 1111
QY 724 AGGAATGTTGTTTATCACTTACAAAGTGGGCTT 759
DB 1112 TGTGCTGTGTTGTTGCTTGTATGCTGTGGGTTT 1147

RESULT 9
ID ABX42390 standard; cDNA; 346 BP.
XX ABX42390;
AC ABX42390;
XX 20-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #7555.
XX
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KM muscle deposition; fat deposition; genome mapping; gene identification;

```

KW gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001; 2001US-0960352.

XX 12-JAN-1999; 99US-115707P.

XX 11-JAN-2000; 2000US-0480902.

XX (BYAT/) BYAT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI, 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat

XX deposition, useful for genome mapping, gene identification and

XX analysis, cattle breeding, or for genetically improving cattle

XX Claim 2: SEQ ID No 7555; 245bp; English.

XX The invention relates to a purified nucleic acid molecule associated with

XX lactation or muscle and fat deposition (designated LMPD), derived

XX from cattle, and the LMPD nucleic acid can specifically hybridise to a

XX second nucleic acid molecule comprising any of 15112 nucleotide

XX sequences, appearing as ABX34836-ABX4947, or complements of them.

XX Also included are: (1) a transformed cell having a nucleic acid

XX comprising an LMPD nucleic acid linked to a promoter and a 3' non-

XX translated sequence that functions in the cell to cause termination of

XX transcription and addition of polyadenylated ribonucleotides to a 3' end

XX of the mRNA molecule; and (2) determining a level or pattern of a

XX molecule in a bovine cell or tissue comprising: (a) incubating a marker

XX nucleic acid (comprising any of the 15112 nucleic acid sequences or its

XX complement or fragment) with a complementary nucleic acid molecule

XX obtained from the bovine cell or tissue, where hybridisation between the

XX marker nucleic acid and the complementary nucleic acid permits the

XX detection of the molecule; and (b) detecting the level or pattern of the

XX complementary nucleic acid, where the detection of the complementary

XX nucleic acid is predictive of the level or pattern of the molecule.

XX The LMPD nucleic acid is used for determining a level or pattern

XX of a molecule in a bovine cell or tissue. It is useful for genome

XX mapping, gene identification and analysis, cattle breeding, preparation

XX of constructs for use in cattle gene expression, or for genetically

XX improving cattle. The present sequence is one of the 15112 bovine

XX LMPD EST (expressed sequence tag) nucleic acids.

XX Note: The present sequence was not shown in the specification but

XX was obtained in electronic format from the USPO web site:

XX segdata.uspto.gov/sequence.html?docID=20020137139..

XX Sequence 346 BP; 96 A; 66 C; 99 G; 85 T; 0 other;

Query Match 25.5%; Score 253; DB 25; Length 346;  
Best Local Similarity 86.8%; Pred. No. 2.5e-61;  
Matches 289; Conservative 0; Mismatches 42; Indels 2; Gaps 1;

```

QY 430 ATATTGAGTAATAGAACTTACGATTCGCTTTCTTACCAAGAACTCCTACTTATGAG 489
DB 14 ATCCGACATATATAGAACTTACGATTCGCTTTCTTACCAAGAACTTATTAATCGAAG 73
QY 430 CACTGGGACTGCTGCCAAGAGCAAGGTGCAAGCTGTTGATGAGAGATTAATACAT 549
DB 74 CACTGGGACTGCTGCCAAGAGCAAGGTGCAAGCTGTTGATGAGAGATTAATACAT 133
QY 550 ATGAGAGAAAGTGGTCAATTAATCTTACTGATGATGATGATGATGATGATG 609
DB 134 ATGAGAGAAAGTGGTCAATTAATCTTACTGATGATGATGATGATGATGATGATG 193

```



QY 610 GGGCTACAGGCTTGTCTGAGTGTGAGAACTGTGCGAGCTGAGAGGGAAGCCGGA 669  
 DB 194 GAGCTACAGGCTTGTCTGAGTGTGAGAACTGTGCGAGCTGAGAGGGAAGCCGGA 253  
 QY 670 AAGAGCAAAAGTCTCTGCTGCAAAAGGTGCTGCGAGCTAATTATTCATGAGGAAC 729  
 DB 254 A--AAGGCAAGTCTCTGCTGCAAAAGGTGCTGCGAGCTAATTATTCATGAGGAAC 311  
 QY 730 TGTGTTGTACACTCTACAGAGTGGGCTTTTC 762  
 DB 312 TGTGTTGTACACTCTACAGAGTGGGCTTTTC 344  
 RESULT 10  
 ABX35482  
 ID ABX35482 standard; cDNA; 314 BP.  
 AC ABX35482;  
 XX  
 DT 20-FEB-2003 (first entry)  
 DE Bovine EST associated with lactation/muscle/fat deposition #647.  
 XX  
 KM Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 KM muscle deposition; fat deposition; genome mapping; gene identification;  
 KM gene analysis; cattle breeding.  
 OS Bos Taurus.  
 XX  
 XX US2002137139-A1.  
 PN 26-SEP-2002.  
 XX  
 PF 24-SEP-2001; 2001US-0960352.  
 XX  
 PR 12-JAN-1999; 99US-115707P.  
 PR 11-JAN-2000; 2000US-0480902.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 XX  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 DR WPI; 2003-110599/10.  
 XX  
 PT New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle -  
 PS  
 PS Claim 2; SEQ ID NO 647; 245bp; English.  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern

CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC [seqdata.uspto.gov/sequence.html?docid=20020137139](http://seqdata.uspto.gov/sequence.html?docid=20020137139).  
 XX  
 SQ Sequence 314 BP; 95 A; 51 C; 91 G; 77 T; 0 other;  
 Query Match 23.9%; Score 238; DB 25; Length 314;  
 Best Local Similarity 86.9%; Pred.No. 4.1e-57;  
 Matches 273; Conservative 0; Mismatches 40; Indels 1; Gaps 1;  
 QY 358 AATGTTGGCTTGTATATGAGTAAGAGCTGCAGAAATATCTATAGAAATCTGCC 417  
 DB 1 AATGTTGGCTTGTATATGAGTAAGAGCTGCAGAAATATCTATAGAAATCTGCC 60  
 QY 418 TGACACCAATGATATGACGTAATAGAACTTCACGATTGCTTCTACCAAGCACTCC 477  
 DB 61 TGACACCAATGATATGACGTAATAGAACTTCACGATTGCTTCTACCAAGCACTTA 120  
 QY 478 TTAATTATGAGCACTGGGACTCTGTCAGAGGCAAGGTCGCAACGCTGTTGATAGAG 537  
 DB 121 TTAATTATGAGCACTGGGACTCTGTCAGAGGTCGCAACGCTGTTGATAGAT 180  
 QY 538 GAGATATATCATATGAGAGAAAGTGGGTCAATTAATCTAGTGGTGAATTTCAAAAG 597  
 DB 181 GAGATATATCATATGAGAGAAAGTGGGTCAATTAATCTAGTGGTGAATTTCAAAAG 240  
 QY 598 GACACCCACTAGGGGCTACAGTCTTGTCTCAGTGTGCAAACTGCTGGCAGC-TGAGA 656  
 DB 241 GGCACCCACTAGGGGCTACAGTCTTGTCTCAGTGTGCAAACTGCTGGCAGCATTGAGA 300  
 QY 657 GGGGAAGCCGGAAA 670  
 DB 301 CGGGGAAGCCGGAAA 314  
 RESULT 11  
 AA199682/c  
 ID AA199682 standard; DNA; 4411529 BP.  
 XX  
 AC AA199682;  
 XX  
 DT 15-JAN-2002 (first entry)  
 XX  
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
 XX  
 KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KM variation; epidemiology; patient treatment; epidemic monitoring; de-  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN US6294328-B1.  
 PN 25-SEP-2001.  
 PD 24-JUN-1998; 98US-0103840.  
 PF 24-JUN-1998; 98US-0103840.  
 PR 24-JUN-1998; 98US-0103840.  
 XX  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
 DR WPI; 2001-647261/74.  
 XX  
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ -

XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

CC The invention relates to evaluating strain variation within and between  
XX different populations of the tuberculosis bacterial pathogen,  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and  
CC H37Rv (A199682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at [seqdata.uspto.gov/Sequence.html?DocID=629432881](http://seqdata.uspto.gov/Sequence.html?DocID=629432881).

XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Query Match 21.5%; Score 214; DB 22; Length 4411529;  
Best Local Similarity 57.1%; Pred. No. 1,9e-48;  
Matches 428; Conservative 0; Mismatches 317; Indels 5; Gaps 2;

QY 6 CCAGTTGCTCTCAGATGTTGGGTATGCTGGAAGAAACATATGGAAGAAATATGAGACA 65

DB 3099716 CCGGTGGCCCGCGATGTTGGGGGGCGCGCGGAAACATAGAAATACGGCACCC 3099657

QY 66 AAAATTGAACCTTGCAAAAATTGGATGAAAATCATTAACATTCAGTTAATACCCG 125

DB 3099656 ACCGCGGACATTTCCGGAAGATCGGCTACAGAACACAGACATCGGTCAACACCCG 3099597

QY 126 TATTCACAGTTCCAGATGAATATACATTTAGATGAAGTACATCTAAGAAAGTTT 185

DB 3099596 TATGACAGATTTAGACAAATACCTTGAGACATCTTGCCCTCAAGATGATTTCC 3099537

QY 186 GATTTTGTACTATCTTACAAATGTTGCCACTTTCAGATGAGTCTGACAGCAATTTG 245

DB 3099536 GACCGGCTACCAAAATGAGAGTCTCCACCTCCGACGGGTGGGGGGGTGGTGGT 3099477

QY 246 GCCAGTGAAGCATTTGTACAGAAATATGAGTCCAAATCCAAAGCTGTGAAATTTTGA 305

DB 3099476 GCCAGTGAAGATTAACCTGCGCAACCAACCTTGGCCGGGCTGTGAAATTCGCGG 3099417

QY 306 CAAGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 365

DB 3099416 CAGGCGATACCAACGATCTGCTTCCACTTTGATG---GCAATGCCGCAATATCATC 3099360

QY 366 GCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 425

DB 3099359 GGTACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3099300

QY 426 AATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 485

DB 3099299 AAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3099240

QY 486 GAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545

DB 3099239 GAGGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3099180

QY 546 ACATATGAGAGAAAGTGTATTAATCTAGTGTGATGATGATGATGATGATGATGATGATG 605

DB 3099179 ACTTACGCGGAGCGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATG 3099120

QY 606 CTAGGCGCTACAGTCTTGTCTAGTGTGAGAACTCTGTGCGAGCTGAGAGGGAAGCC 665

DB 3099119 CTGGGCGCGAGCGGATGCGGAGTGTGCGGAGTACCTGTGCGAGTGTGCGGACCGCC 3099060

QY 666 GGAAGAGAGGCAAGTCTCTGTGTGCAAGAGTGTGCTGTGCGCATTAATTANGATTGAG 725

DB 3099059 G--AGCGCGCGCAGGTGCAACAGTGAACCGCGCGCTGCAACACATCGGAGTGGGG 3099002

QY 726 GAAGTGTGTGTTAACAATCTACAGATG 755

DB 3099001 GCGCCGCGGTGTACCGGATACCAAGCG 3098972

RESULT 12

AB16660/C

ID AB16660 standard; DNA; 3613 BP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Claim 1; SEQ ID NO 1453; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB16511), expressed DNA  
CC sequences (AB161840-AB16175) and the encoded proteins  
CC (AB161737-AB162072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_ptc\\_sequences](http://ftp.wipo.int/pub/published_ptc_sequences).

XX Sequence 3613 BP; 948 A; 876 C; 843 G; 946 T; 0 other;

Query Match 21.3%; Score 211.4; DB 23; Length 3613;  
Best Local Similarity 60.3%; Pred. No. 4.1e-49;  
Matches 384; Conservative 0; Mismatches 248; Indels 5; Gaps 2;

QY 126 TATTCAGATTCAGATGAATACAGTTTAATGATGATGATGATGATGATGATGATGATG 185

DB 1725 TACTGCGATTCGCGAGTGAATACACCTGAGAGATTAATGAAGTCCGCCAGGTGTG 1666

QY 186 GA---TTTGTGATATCTTACATATGTTGCTCCCTTCCAGATGATGATGATGATGATG 242

DB 1665 GAGGAGTGTGACCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1606

QY 243 TTGGCAGTGAAGCATTTGTATCAGAAATGATGATGATGATGATGATGATGATGATG 302

DB 1605 CTCGCTCCGAGGCTTGTGTGCTGCGCCAGATTTGAGAAAGCAGCTGTGTGAGATTGTG 1546

QY 303 GCACAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 362

DB 1545 GGCATGAGATGCGCAGTGAACCGGCGCTCCACTTTGCGGACAAAGAGCTGATGAAGATT 1486





Oy	303	GCACAGAAATGATGACTGATTTGCGAAAGCTGTTTGAAAGAAAAAGATATATTAATG	362
Db	2317	GGCATGAGATGCGCAGTACCCGGCGTCCACTTTGGCCAGAAAGCTTGATBAAGTT	2376
Oy	363	GTTGGCTTTGATATGAGTAAAGAGCTGCAGAAAAATGCTATYGAAATCTGGCTGACA	422
Db	2377	GCTGGAAACCGATATGACCCGCTGTGGCCACCAGCGTCTGTTCGCAAAAGAGGGTACAAG	2436
Oy	423	CCAAATGATATTGACGTAAATGAACTTCACAGTTGCTTTTCTACCAAGAACTCCTTACT	482
Db	2437	CCCCAGAGTGTCCAGGTGTGGAACCTGACGATTTCTTCGGCCAAACGATTTATACG	2496
Oy	483	TATGAAGCACTGGGACTCTGTCCAGAAAGCAAGGTGCACGCTGTTGATAGAGAGAT	542
Db	2497	TATAGGCACTTGGACTGTGTGGAGAGGGCAACCGCGAGTTTCATGACGCTGGAGAC	2556
Oy	543	AATACATATGAGAGAAAGTGGCTATTAATCTTACTGTGGACTGATTTCAAAGGACAC	602
Db	2557	AACACCTAACGGTGGCAAGTTGTGTGTCAACCCAGTGTGTCTGATCTCCAAAGGCGAT	2616
Oy	603	CCACTAGGCGCTAACAGTCTTGCTCAGTGTCCAGAACTCTGTGGACGTGAGAGGGAA	662
Db	2617	CTCTTGGGGCCGACGGGTCTGGCACAAATGTCTGAAGCTCTGTGGCACTCCGTGAAATG	2676
Oy	663	GCCGGAAGAGGCAAAAGTTCTTGGTGCMAAGGTGGCTCTGCACGATTAATTANGACTTG	722
Db	2677	GCTG-AGAAACGGCCAGGTGCCAATGCTCAGTTGGCTTTCCACACACATCTGGAGACTGG	2734
Oy	723	GAGAACGTGTGTTGTAACTCTCAACAAGATGGGGTT	759
Db	2735	GTGTGCGCGTGTGTGGTCTTGTATCTCTGGGTTT	2771

Search completed: November 27, 2003, 06:23:51  
Job time : 334.626 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 06:06:02 ; Search time 374.822 Seconds

(without alignments)  
8737.270 Million cell updates/sec

Title: US-09-835-992A-21

Perfect score: 994  
Sequence: 1 ctaccaccgtctgcctccacg.....ccaccctgggtgtgtgggat 994

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2180069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	978	98.4	994	US-09-835-992A-21	Sequence 21, Appl
2	322.8	33.5	432	US-09-960-352-5678	Sequence 5678, Ap
3	253	25.5	346	US-09-960-352-7555	Sequence 7555, Ap
4	238	23.9	314	US-09-960-352-647	Sequence 647, Ap
5	196.2	13.7	1236	US-10-156-761-5843	Sequence 5843, Ap
6	196.2	13.7	902560	US-10-156-761-1	Sequence 1, Appl
7	189.8	15.1	412	US-09-960-352-2946	Sequence 2946, Ap
8	92	9.3	249	US-09-878-574-8890	Sequence 8890, Ap
9	89.4	9.0	553	US-10-102-524-364	Sequence 364, App
10	67.2	6.8	938	US-10-027-632-34138	Sequence 34138, A
11	67.2	6.8	938	US-10-027-632-34138	Sequence 34138, A
12	67.2	6.8	938	US-10-027-632-34138	Sequence 34138, A
13	67.2	6.8	938	US-10-027-632-34138	Sequence 34138, A
14	42.6	4.3	2432	US-10-027-632-103694	Sequence 103694, A
15	42.6	4.3	2432	US-10-027-632-103694	Sequence 103694, A
16	42.6	4.3	2432	US-10-027-632-103694	Sequence 103694, A
17	42.6	4.3	2432	US-10-027-632-103694	Sequence 103694, A

C 17	42.6	4.3	2432	13	US-10-027-632-103694	Sequence 103694,
C 18	42.6	4.3	2432	13	US-10-027-632-103695	Sequence 103695,
C 19	42.6	4.3	2432	13	US-10-027-632-111751	Sequence 111751,
C 20	40.2	4.0	3673778	12	US-10-312-841-2	Sequence 2, Appl1
C 21	39.8	4.0	6313	8	US-08-781-986A-201	Sequence 201, App
C 22	39.4	4.0	671	14	US-10-184-644-146	Sequence 346, App
C 23	39.4	4.0	671	14	US-10-184-644-146	Sequence 346, App
C 24	39.4	4.0	580073	12	US-10-205-220-1	Sequence 1, Appl1
C 25	38.6	3.9	9207	12	US-10-240-453-307	Sequence 307, App
C 26	38.2	3.8	1659	9	US-09-815-242-4442	Sequence 4442, Ap
C 27	38.2	3.8	1659	9	US-09-815-242-8331	Sequence 8331, Ap
C 28	38	3.8	1161	12	US-10-027-632-260949	Sequence 260949,
C 29	38	3.8	1161	13	US-10-027-632-260949	Sequence 260949,
C 30	38	3.8	10279	12	US-10-311-455-1563	Sequence 1563, Ap
C 31	37.6	3.8	3673778	12	US-10-312-841-1	Sequence 1, Appl1
C 32	37.4	3.8	494	12	US-10-029-386-20259	Sequence 20259, A
C 33	37.4	3.8	1269	10	US-09-881-752A-91	Sequence 91, Appl1
C 34	37.2	3.7	1161	12	US-10-027-632-260946	Sequence 260946,
C 35	37.2	3.7	1161	12	US-10-027-632-260947	Sequence 260947,
C 36	37.2	3.7	1161	12	US-10-027-632-260948	Sequence 260948,
C 37	37.2	3.7	1161	13	US-10-027-632-260946	Sequence 260946,
C 38	37.2	3.7	1161	13	US-10-027-632-260947	Sequence 260947,
C 39	37.2	3.7	1161	13	US-10-027-632-260948	Sequence 260948,
C 40	37.2	3.7	640681	10	US-09-790-988-1	Sequence 1, Appl1
C 41	37	3.7	810	12	US-10-027-632-159529	Sequence 159529,
C 42	37	3.7	810	12	US-10-027-632-159530	Sequence 159530,
C 43	37	3.7	810	13	US-10-027-632-159529	Sequence 159529,
C 44	37	3.7	810	13	US-10-027-632-159530	Sequence 159530,
C 45	37	3.7	7589	12	US-10-240-453-264	Sequence 264, App

#### ALIGNMENTS

RESULT 1  
US-09-835-992A-21  
Sequence 21, Application US/09835992A  
Patent No. US20020037541A1  
GENERAL INFORMATION:  
APPLICANT: Obata, Yuichi  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND TREATING GASTRIC CANCER  
FILE REFERENCE: L0461/7112  
CURRENT FILING DATE: 2001-04-16  
PRIOR APPLICATION NUMBER: US 08/896,164  
PRIOR FILING DATE: 1997-07-17  
NUMBER OF SEQ ID NOS: 87  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 21  
LENGTH: 994  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Unsure  
LOCATION: (705)..(705)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (716)..(716)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (807)..(807)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (830)..(830)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (833)..(833)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (841)..(841)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure

```
LOCATION: (843) .. (843)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (851) .. (851)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (862) .. (862)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (879) .. (879)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (881) .. (881)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (892) .. (892)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (918) .. (918)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (922) .. (922)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (949) .. (949)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (972) .. (972)
OTHER INFORMATION: n = a, c, g or t
US-09-835-992a-21
```

Query Match 98.4%; Score 978; DB 9; Length 994;

Best Local Similarity 100.0%; Pred. No. 2,9e-248; Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTCACCCAGTGTGCTCTCAGATGTTGGGATGCTGGAAGAACATATGAAAATATG 60
DB 1 CTCACCCAGTGTGCTCTCAGATGTTGGGATGCTGGAAGAACATATGAAAATATG 60
QY 61 GAACAAAAATGTAACATTTGGCAAAATTTGATGAAAAATCATTAACATTCAGTTAATA 120
DB 61 GAACAAAAATGTAACATTTGGCAAAATTTGATGAAAAATCATTAACATTCAGTTAATA 120
QY 121 ACCCGATTCCCAAGTCCCAAGATGATTAAGTTAATGAAGTATGCGATCTTAAGAAG 180
DB 121 ACCCGATTCCCAAGTCCCAAGATGATTAAGTTAATGAAGTATGCGATCTTAAGAAG 180
QY 181 TTTTGGATTTTGGATCTATCTTACATGTTGTCCTCCACTTCAGATGCTGCAGACAA 240
DB 181 TTTTGGATTTTGGATCTATCTTACATGTTGTCCTCCACTTCAGATGCTGCAGACAA 240
QY 241 TTTTGGCCAGTGAAGCATTTGTACAGAAATGAGCTGCAATTCAAAGCTGTGAATTT 300
DB 241 TTTTGGCCAGTGAAGCATTTGTACAGAAATGAGCTGCAATTCAAAGCTGTGAATTT 300
QY 301 TGGCCCAAAATGATGATGATTTGGCCCAAGCTCGTTTGAAGAAAAACATATTAATA 360
DB 301 TGGCCCAAAATGATGATGATTTGGCCCAAGCTCGTTTGAAGAAAAACATATTAATA 360
QY 361 TGGTGGCTTTGATATGATGAAGAAAGCTGCAAGAAATGCTATGAGAATCTGGCTGA 420
DB 361 TGGTGGCTTTGATATGATGAAGAAAGCTGCAAGAAATGCTATGAGAATCTGGCTGA 420
QY 421 CACCAAAATGATATGATGAAGAAATGCAAGATGCTTTTCTACCAAGAACTCTCTTA 480
DB 421 CACCAAAATGATATGATGAAGAAATGCAAGATGCTTTTCTACCAAGAACTCTCTTA 480
QY 481 CTTAAGAACACTGGAAGCTGCTGCGAAGAGACAAAGTCAACGCTGTTGATAGAGAG 540
DB 481 CTTAAGAACACTGGAAGCTGCTGCGAAGAGACAAAGTCAACGCTGTTGATAGAGAG 540
QY 541 ATAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGGAGCTGATTTCAAGAGGAG 600
DB 541 ATAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGGAGCTGATTTCAAGAGGAG 600
```

```
DB 541 ATAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGGAGCTGATTTCAAGAGGAG 600
QY 601 ACCCACTGAGGCTTACAGGTCTTGTCTCAGTGTGCAAACTGTCTGGCAAGTGAAGGGG 660
DB 601 ACCCACTGAGGCTTACAGGTCTTGTCTCAGTGTGCAAACTGTCTGGCAAGTGAAGGGG 660
QY 661 AAGCCGAAAAGAGCAAAAGTTCCTGTGCAAAAGTGGCTGTGCGCATTAATTTANGAT 720
DB 661 AAGCCGAAAAGAGCAAAAGTTCCTGTGCAAAAGTGGCTGTGCGCATTAATTTANGAT 720
QY 721 TGGAGAACTGTGATTTGTAACACTCTACAAATGAGGATTTTCCGAAACCGCAGTTCC 780
DB 721 TGGAGAACTGTGATTTGTAACACTCTACAAATGAGGATTTTCCGAAACCGCAGTTCC 780
QY 781 TTTTGAAGCTCATCAAAATTTGAAGCCTGTTCCAAACAGCTCTGCAAGTATNGTTTAA 840
DB 781 TTTTGAAGCTCATCAAAATTTGAAGCCTGTTCCAAACAGCTCTGCAAGTATNGTTTAA 840
QY 841 NNNAAATCTGTTTAAAGNGGATGAGAAAGAAACNTNAAAGAGGAAAGGGAACA 900
DB 841 NNNAAATCTGTTTAAAGNGGATGAGAAAGAAACNTNAAAGAGGAAAGGGAACA 900
QY 901 ATTTGTGAAGAAATATGCGNGGGAATTTTTCCTTCAAGGGAAATATGAGCTGG 960
DB 901 ATTTGTGAAGAAATATGCGNGGGAATTTTTCCTTCAAGGGAAATATGAGCTGG 960
QY 961 GGGTAAAGAAAGGCGCACTTGGGGTGGGAT 994
DB 961 GGGTAAAGAAAGGCGCACTTGGGGTGGGAT 994
```

## RESULT 2

US-09-960-352-5678

; Sequence 5678; Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengping

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 5678

; LENGTH: 432

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 24-LIB34-074-Q1-E1-F7

US-09-960-352-5678

Query Match 32.5%; Score 322.8; DB 10; Length 432;

Best Local Similarity 88.7%; Pred. No. 4e-75; Matches 360; Conservative 0; Mismatches 44; Indels 2; Gaps 1;

```
QY 357 AAAATGTTGGCTTGTATATGATGAAGAGCTGCAAGAAATGCTATGAGAAATCTGGC 416
DB 2 AAAATGTTGGCTTGTATATGATGAAGAGCTGCAAGAAATGCTATGAGAAATCTGGC 61
QY 417 CTGAACCAAAATGATATGAGCTATTAAGAACTTCCAGTTGTTTCTACCAAGAACTC 476
DB 62 CTGAACCAAAATGATATGAGCTATTAAGAACTTCCAGTTGTTTCTACCAAGAACTC 121
QY 477 CTTACTTAAAGAACTGGAAGCTGTCTCAGAGAGCAAGGTCAACCTGGTGTATAGA 536
DB 122 ATTACTTAAAGAACTGGAAGCTGTCTCAGAGAGCTCAAGGTGAAACTGGTTAAAGA 181
QY 537 GGAAGTAAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGTGAGCTGATTTCAAG 596
DB 182 GGAAGTAAATACATATGAGAGAAAGTGGTCAATTAATCTAGTGTGAGCTGATTTCAAG 241
```



RESULT 4  
US-09-960-352-647  
Sequence 647, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengping  
APPLICANT: Byatt, John C.  
APPLICANT: Machialago, Nagappan

```

/      TYPE: DNA
/      ORGANISM: Streptomyces avermitilis
/      FEATURE:
/      NAME/KEY: CDS
/      LOCATION: (1)..(1236)
/      US-10-156-761-5843
Query Match      19.7%; Score 196.2; DB 14; Length 1236;

```

Best Local Similarity 55.1%; Pred. No. 2,2e-41;  
Matches 402; Conservative 0; Mismatches 325; Indels 2; Gaps 1;

```

QY 16 CTCAGATGTTGGGATGCTGGAAGAAACATATGGAAGAAATATGGAACAAATTTGAAC 75
   |||||
DB 476 CCAGATCTTCGGCGACGCGCGCGGACACATGAGAAATGCGGAGACGACCGAGCGCG 535
   |||||
QY 76 ACTTGGCAAAATATGATGAGAAATCATATAATTCAGTTAATACCGGTATTCGCAAT 135
   |||||
DB 536 AACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 595
   |||||
QY 136 TCCAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 195
   |||||
DB 596 TCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 655
   |||||
QY 196 CTATCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255
   |||||
DB 656 CCAAGCTCCAGTGTCTCAGCCACCTCGGACGAGCGCGCGCGCGCGCGCGCGCGCGCG 715
   |||||
QY 256 CATTTGTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 315
   |||||
DB 716 GCTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 775
   |||||
QY 316 TGAATGATTTGCCAAGCTCGTTTGAAGAAATATTAATGATGATGATGATGATGATGATGAT 375
   |||||
DB 776 CCAAGCAACGAGAGAGATCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 835
   |||||
QY 376 TGAATGAAAGAGCTGCAAGAAATGCTATGAGAAATCTGCGCTGACACCAATGATATTTG 435
   |||||
DB 836 TGTTCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
   |||||
QY 436 AGCTAATGAACTTCAAGATGCTTTTCAACCAAGCACTCTTACTTATGAGCACTGG 495
   |||||
DB 896 AGCTGATGAACTGCAAGATGCTTTTCAACCAAGCACTCTTACTTATGAGCACTGG 955
   |||||
QY 496 GACTCTGTCCAGAGAGCAAGAGTGCACAGCTGCTGATGAGAGAGATTAATCAATATGAG 555
   |||||
DB 956 GCATGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
   |||||
QY 556 GAAAGTGGGTCAATAATCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 615
   |||||
DB 1016 GCCGGTGGGTGCGGAAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
   |||||
QY 616 CAGGTCTTCTCAGTGTGCAAGACTCTGCTGCGAGCTGAGAGGGAAGCCGGAAGAGAG 675
   |||||
DB 1076 CGGATATGCGCGAGATGCGCGAGCTGACCTGCGAGCTGCGCGCGCGCGCGCGCGCGCG 1133
   |||||
QY 676 CAAAGTCTCTGTCGCAAGAGTGGCTCTGCGAGCATTAATTANGCATTTGAGAGAACTGTGT 735
   |||||
DB 1134 CCAAGTCTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1193
   |||||
QY 736 TGTAACT 744
   |||||
DB 1194 GGTGACGCT 1202
   |||||

```

RESULT 6  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 19.7%; Score 196.2; DB 14; Length 9025608;  
Best Local Similarity 55.1%; Pred. No. 3e-39;  
Matches 402; Conservative 0; Mismatches 325; Indels 2; Gaps 1;

```

QY 16 CTCAGATGTTGGGATGCTGGAAGAAACATATGGAAGAAATATGGAACAAATTTGAAC 75
   |||||
DB 7075437 CGAGATCTTCGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7075496
   |||||
QY 76 ACTTGGCAAAATATGATGAGAAATCATATAATTCAGTTAATACCGGTATTCGCAAT 135
   |||||
DB 7075497 AACTGCGCGCGGTGCGGGGCAAGAACACCGGCACTCGGCGCGCAACCGGTAGCGCGAGT 7075556
   |||||
QY 136 TCCAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 195
   |||||
DB 7075557 TCCAGAGAGCTGATACCGGTGACGAGATCTTCGCGCGCGCGCGCGCGCGCGCGCG 7075616
   |||||
QY 196 CTATCTTCAATGTTGTCCCACTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 255
   |||||
DB 7075617 CCAAGCTCAGTGTCTCAACCTGACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7075676
   |||||
QY 256 CATTTGTACAGAGATGAGCTGCAATCCAAAGCTGTGGAATTTTGGCACAAGAAATGA 315
   |||||
DB 7075677 GCTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7075736
   |||||
QY 316 TGAATGATTTGCCAAGCTCGTTTGAAGAAATTAATTAATGATGATGATGATGATGATGATGAT 375
   |||||
DB 7075737 GCACGGAACGAGAGAGATCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7075796
   |||||
QY 376 TGAATGAAAGAGCTGCAAGAAATGCTATGAGAAATCTGCGCTGACACCAATGATATTTG 435
   |||||
DB 7075797 TGTTCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7075856
   |||||
QY 436 AGCTAATGAACTTCAAGATGCTTTTCAACCAAGCACTCTTACTTATGAGCACTGG 495
   |||||
DB 7075857 AGCTGATGAACTGCAAGATGCTTTTCAACCAAGCACTCTTACTTATGAGCACTGG 7075916
   |||||
QY 496 GACTCTGTCCAGAGAGCAAGTGCACAGCTGCTGATGAGAGAGATTAATCAATATGAGAG 555
   |||||
DB 7075917 GCATGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7075976
   |||||
QY 556 GAAAGTGGGTCAATAATCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 615
   |||||
DB 7075977 GCCGTGGGTGCTGAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7076036
   |||||
QY 616 CAGGTCTTCTCAGTGTGCAAGACTCTGCTGCGAGCTGAGAGGGAAGCCGGAAGAGAG 675
   |||||
DB 7076037 CGGATATGCGCGAGATGCGCGAGCTGACCTGCGAGCTGCGCGCGCGCGCGCGCGCGCG 7076094
   |||||
QY 676 CAAAGTCTCTGTCGCAAGAGTGGCTCTGCGAGCATTAATTANGCATTTGAGAGAACTGTGT 735
   |||||
DB 7076095 CCAAGTCTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7076154
   |||||
QY 736 TGTAACT 744
   |||||
DB 7076155 GGTGACGCT 7076163
   |||||

```

RESULT 7  
US-09-960-352-2946



```
US-10-027-632-34138/c
; Sequence 34138, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIORITY FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34138
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(938)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-34138

Query Match
Best Local Similarity 6.8%; Score 67.2; DB 12; Length 938;
Matches 75; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 209 TTGTCCCACTTGCAGATGCTGCTGCAGCAGCAATTTTGGCCAGTGAAGCATTGTACAGAA 268
DB 652 TAGTCTTACTCAGATGCTGCTGCAGTGCAGCAATTTTGGCCAGTGAAGCATTGTACAGAA 593
QY 269 GTATGCGCTGCAATCCAAAGCTGTGGAAT 298
DB 592 GTATGCGCTGCAACCAAGAGTTCCACAT 563

RESULT 11
US-10-027-632-34139/c
; Sequence 34139, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIORITY FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
```

```
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34139
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(938)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-34139

Query Match
Best Local Similarity 6.8%; Score 67.2; DB 12; Length 938;
Matches 75; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 209 TTGTCCCACTTGCAGATGCTGCTGCAGCAGCAATTTTGGCCAGTGAAGCATTGTACAGAA 268
DB 652 TAGTCTTACTCAGATGCTGCTGCAGTGCAGCAATTTTGGCCAGTGAAGCATTGTACAGAA 593
QY 269 GTATGCGCTGCAATCCAAAGCTGTGGAAT 298
DB 592 GTATGCGCTGCAACCAAGAGTTCCACAT 563

RESULT 12
US-10-027-632-34138/c
; Sequence 34138, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIORITY FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34138
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(938)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-34138

Query Match
Best Local Similarity 6.8%; Score 67.2; DB 13; Length 938;
Matches 75; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 209 TTGTCCCACTTGCAGATGCTGCTGCAGCAGCAATTTTGGCCAGTGAAGCATTGTACAGAA 268
DB 652 TAGTCTTACTCAGATGCTGCTGCAGTGCAGCAATTTTGGCCAGTGAAGCATTGTACAGAA 593
QY 269 GTATGCGCTGCAATCCAAAGCTGTGGAAT 298
DB 592 GTATGCGCTGCAACCAAGAGTTCCACAT 563
```

```
RESULT 13
US-10-027-632-34139/c
; Sequence 34139, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34139
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(938)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-34139

Query Match      6.8%; Score 67.2; DB 13; Length 938;
Best Local Similarity 83.3%; Pred. No. 2.5e-07;
Matches 75; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Qy      209 TTGTCACCTTCAGATGGTGCTGTCAGACGACATTTTGGCCAGTGAAGCATTTGTACAGA 268
Db      652 TAGTCTTACTCTAGATGCTGCTGCTGCTGCAATTTTGGCCAGKGAAGCATTTTGCACAGA 593
Qy      269 GTATGGCTTCGACATCAAAAGCTGTGGAAT 298
Db      592 GTATGGCTTCGACAAACCAAGAGTTCACAT 563

RESULT 14
US-10-027-632-103694/c
; Sequence 103694, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
```

```
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103694
; LENGTH: 2432
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-103694

Query Match      4.3%; Score 42.6; DB 12; Length 2432;
Best Local Similarity 51.9%; Pred. No. 1.4;
Matches 96; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy      305 ACAAGAAATGATGACTGATTTTGGCAAGCTCGTTGAGAAAAAGCATTTATTAATGCT 364
Db      416 ACATGAAAAATATATCTAGCTTAAATAGATTAATTGAATMAAAAGCTTTTCAGATATGCT 357
Qy      365 TGGCTTGATATGAGTAAGAAGCTGCAAGAAAATGCTATGGAATCTGGCTGACACC 424
Db      356 TTTCATGTAGATATCACAAGCATCTGCCAGAGTAAGCAATTAATTTCTTTGAAAT 297
Qy      425 AATGATATTTGACGTAATAGAACTTCAGATTTGCTTTCTACCAAGAACTCTTACTTA 484
Db      296 TAAAAAGGAAACGAAAAATTCATATCTTTTTTTATTTTCCAAAGCTTCACAGA 237
Qy      485 TGAG 489
Db      236 AGAAG 232
```

```
RESULT 15
US-10-027-632-103695/c
; Sequence 103695, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103695
; LENGTH: 2432
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-103695

Query Match      4.3%; Score 42.6; DB 12; Length 2432;
Best Local Similarity 51.9%; Pred. No. 1.4;
Matches 96; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy      305 ACAAGAAATGATGACTGATTTTGGCAAGCTCGTTGAGAAAAAGCATTTATTAATGCT 364
Db      416 ACATGAAAAATATCTAGCTTAAATAGATTAATTGAATMAAAAGCTTTTCAGATATGCT 357
```

QY	365	TGGCTTTGATGATGATGAAGAACTGCAAGAAAATGCTATGAGAAATCTGGCCTGACACC	424
Db	356	TTTCATGTAAGTATCACAGCACTGCGAGAGTAAGCAATTAGTATTCTTTGAAAT	297
QY	425	AAATGATATTGACGTAAATAGAACTTCACGATTGCTTTCTACCAAGAACTCCTTACTTA	484
Db	236	TAAAAAGGAAACAGAAAAATCATTAATTTTTTTTATTTTCCAACTGCTTCAGAA	237
QY	485	TGAAG	489
Db	236	AGAAG	232

Search completed: November 27, 2003, 12:53:07  
Job time : 384.822 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 06:00:22 ; Search time 73.1681 Seconds  
(without alignments) 5797.910 Million cell updates/sec

Title: US-09-835-992A-21

Perfect score: 994  
Sequence: 1 CTCACCCAGTCGCTCCTCAGATGTTGGCGTATGCTGAAAAAGACATATGCAAAATATG 994

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	978	98.4	994	3	US-08-896-164-21
2	214	21.5	441529	3	US-09-103-840A-1
3	199.6	20.1	4403765	3	US-09-103-840A-2
4	92	9.3	1664976	4	US-08-916-421B-1
5	49.2	4.9	7218	1	US-08-232-463-14
6	39.8	4.0	1644	1	US-08-785-048-4
7	39.8	4.0	1644	2	US-08-996-799-4
8	39.8	4.0	1662	1	US-08-785-048-1
9	39.8	4.0	1662	2	US-08-996-799-1
10	39	3.9	580073	4	US-08-545-528D-1
11	38.2	3.8	1368	4	US-09-328-352-1034
12	37.8	3.8	1664976	4	US-08-916-421B-1
13	37.4	3.8	1353	4	US-09-252-991A-13040
14	37.4	3.8	1410	2	US-09-252-991A-12438
15	35.8	3.6	19124	2	US-08-487-825B-13
16	35.8	3.6	1830121	4	US-09-557-884-1
17	35.8	3.6	1830121	4	US-09-643-990A-1
18	35	3.5	1464	3	US-08-605-150A-3
19	34.8	3.5	393	4	US-09-328-352-826
20	34.8	3.5	2079	4	US-09-381-849-4
21	34.8	3.5	2193	4	US-09-328-352-316
22	34.6	3.5	1301	1	US-08-431-080-1
23	34.6	3.5	1301	1	US-08-431-080-4
24	34.6	3.5	1301	2	US-08-938-534-1
25	34.6	3.5	1301	2	US-08-938-534-4
26	34.6	3.5	1301	4	US-09-345-294-1
27	34.6	3.5	1301	4	US-09-345-294-4

28	34.6	3.5	1490	3	US-08-605-150A-5	Sequence 5, Appl1
29	34.4	3.5	580073	4	US-08-545-528D-1	Sequence 1, Appl1
30	34.2	3.4	599	3	US-09-328-111-219	Sequence 219, App
31	33.8	3.4	572	3	US-09-328-111-87	Sequence 87, App
32	33.8	3.4	614	3	US-09-328-111-111	Sequence 111, App
33	33.8	3.4	10411	4	US-08-961-527-89	Sequence 89, Appl
34	33.6	3.4	22067	4	US-09-820-001-3	Sequence 3, Appl1
35	33.6	3.4	37948	3	US-09-251-645-11	Sequence 11, Appl
36	33.4	3.4	3663	3	US-09-499-884-11	Sequence 11, Appl
37	33.4	3.4	15016	4	US-09-601-198-60	Sequence 60, Appl
38	33.2	3.3	582	4	US-09-107-532A-1201	Sequence 1201, Ap
39	33.2	3.3	3627	1	US-08-447-500-23	Sequence 23, Appl
40	33.2	3.3	3627	1	US-08-454-097-23	Sequence 23, Appl
41	33.2	3.3	3627	1	US-08-453-866-23	Sequence 23, Appl
42	33.2	3.3	3627	1	US-08-185-359-23	Sequence 23, Appl
43	33.2	3.3	3628	1	US-08-468-016-17	Sequence 17, Appl
44	33.2	3.3	3628	2	US-08-376-843-17	Sequence 17, Appl
45	33.2	3.3	9873	4	US-09-328-352-1360	Sequence 1360, Ap

## ALIGNMENTS

```
RESULT 1
US-08-896-164-21
; Sequence 21, Application US/08896164
; Patent No. 6218521
; GENERAL INFORMATION:
; APPLICANT: OBATA, Yuch1
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR
; TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/896,164
; FILING DATE: July 17, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6218521man D. Hanson
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-896-164-21
;
Query Match          98.4%; Score 978; DB 3; Length 994;
Best Local Similarity 100.0%; Pred. No. 2.2e-271;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 CTCACCCAGTCGCTCCTCAGATGTTGGCGTATGCTGAAAAAGACATATGCAAAATATG 60
        1 CTCACCCAGTCGCTCCTCAGATGTTGGCGTATGCTGAAAAAGACATATGCAAAATATG 60
Db       1 CTCACCCAGTCGCTCCTCAGATGTTGGCGTATGCTGAAAAAGACATATGCAAAATATG 60
QY      61 GAACAAAATTGAACACTTTCAGAAAATTGGATGAGAAAATCATTAACATTAGTTAATA 120
```

```

Db      61  GAAACAAAATTGAACCTTTGCAAAAATTGATGAGAAAATCATTAACATTCCTGTTAA 120
Qy      121  ACCCGATTCCCGAGTTCCAGATGAATACGTTTAGATGAAGTGGCATCTAAAGAG 180
Db      121  ACCCGATTCCCGAGTTCCAGATGAATACGTTTAGATGAAGTGGCATCTAAAGAG 180
Qy      181  TTTTGAATTTTGAATCTTACATGTTTGTCCCATTCAGATGTTGCTGCGACGCA 240
Db      181  TTTTGAATTTTGAATCTTACATGTTTGTCCCATTCAGATGTTGCTGCGACGCA 240
Qy      241  TTTTGGCCAGTGAAGATTTGTACAGAAATGAGTGGCTGCAATCCAAAGCTGGAATTT 300
Db      241  TTTTGGCCAGTGAAGATTTGTACAGAAATGAGTGGCTGCAATCCAAAGCTGGAATTT 300
Qy      301  TGGCACAAGAAATGATGACTGATTTTCCAGCTCGTTTGAAGAAAAAGCATTTTAA 360
Db      301  TGGCACAAGAAATGATGACTGATTTTCCAGCTCGTTTGAAGAAAAAGCATTTTAA 360
Qy      361  TGGTGGCTTTGATATGAGTAAAGAGCTGCAAGAAATGCTATGGAATCTGGCTGA 420
Db      361  TGGTGGCTTTGATATGAGTAAAGAGCTGCAAGAAATGCTATGGAATCTGGCTGA 420
Qy      421  CACCAATGATTTGACGTATAGAACTTCAAGATGCTTTTCTACCAAGAACTCCTTA 480
Db      421  CACCAATGATTTGACGTATAGAACTTCAAGATGCTTTTCTACCAAGAACTCCTTA 480
Qy      481  CTTTGAAGCACTGGGACTCTGTCCAGAAAGACAAAGTGCACGCTGTTGATAGAG 540
Db      481  CTTTGAAGCACTGGGACTCTGTCCAGAAAGACAAAGTGCACGCTGTTGATAGAG 540
Qy      541  ATTAATCATATGAGAAAGTGGTCTAATCTCTAGTGTGATCTGATTTCAAAGGAC 600
Db      541  ATTAATCATATGAGAAAGTGGTCTAATCTCTAGTGTGATCTGATTTCAAAGGAC 600
Qy      601  ACCCACTAGGCGCTCAGGCTCTGTCAAGTGCAGAACTCTGCTGGCAGCTGAGAG 660
Db      601  ACCCACTAGGCGCTCAGGCTCTGTCAAGTGCAGAACTCTGCTGGCAGCTGAGAG 660
Qy      661  AAGCCGGAAGAGGCAAGATCTCTGTCAAGGATGGCTCTGCGAAGCCGCACTTCC 720
Db      661  AAGCCGGAAGAGGCAAGATCTCTGTCAAGGATGGCTCTGCGAAGCCGCACTTCC 720
Qy      721  TGAAGAACTGTGCTGTGAACCTTCAAGATGGGGTTTCCCGAAGCCGCACTTCC 780
Db      721  TGAAGAACTGTGCTGTGAACCTTCAAGATGGGGTTTCCCGAAGCCGCACTTCC 780
Qy      781  TTTTGAACCTCATCAAAATTTGAAGCNGTTCCAAAGCTCTGCAAGTATNGTTTAA 840
Db      781  TTTTGAACCTCATCAAAATTTGAAGCNGTTCCAAAGCTCTGCAAGTATNGTTTAA 840
Qy      841  NGNAATATCTNGTTTAAAGNGAGATTGAGAAAGAACTNNAAGAGGAGAGGAGACA 900
Db      841  NGNAATATCTNGTTTAAAGNGAGATTGAGAAAGAACTNNAAGAGGAGAGGAGACA 900
Qy      901  ATTTTGAAGAAATTTGNGGNGGAAATTTTTCCTTCAAGGGGAAATATGCCCCCTGG 960
Db      901  ATTTTGAAGAAATTTGNGGNGGAAATTTTTCCTTCAAGGGGAAATATGCCCCCTGG 960
Qy      961  GGGGTAAAGAGGCAAGCTGGGGTGTGGGAT 994
Db      961  GGGGTAAAGAGGCAAGCTGGGGTGTGGGAT 994

```

## RESULT 2

```

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.

```

```

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match      21.5%; Score 214; DB 3; Length 4411529;
Best Local Similarity 57.1%; Pred. No. 3; 7e-50;
Matches 428; Conservative 0; Mismatches 317; Indels 5; Gaps 2;

```

```

Qy      6  CCAATTGCTCTCCATGATGTTGGGTATGCTGGAAGAAACATATGGAATAATGAGACA 65
Db      3099716  CCGTGGCGCGGTGATGTTGGGGGCGCGCGCGGCAACATGAAAGAAATGCGCAC 3099657
Qy      66  AAAATTGAACATTTGCAAAAATTGATGAGAAATAATCAATTCAGTTAATACCG 125
Db      3099656  ACCGGGAGCATTTGCGGAAGATGGCTTAAGAAACCAAGCACTCGGTCAACACCG 3099597
Qy      126  TATTCACAGTTCCAAAGATGATACGTTTATGATGATGATGATCTTAAGAGTTT 185
Db      3099596  TATGACAGATTTCAAGAGCAATACACCTGACGACATCTTGGCTCAAGATGATTTCC 3099537
Qy      186  GATTTTGTGATCTTCAATGTTGTCCTCCATTCAGATGATGATGATGATGATGAT 245
Db      3099536  GACCGCTGACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3099477
Qy      246  GCCAGTGAAGCATTTGTACAGAAATGATGATGATGATGATGATGATGATGATGAT 305
Db      3099476  GCCAGTGAAGCATTTGTACAGAAATGATGATGATGATGATGATGATGATGATGAT 3099417
Qy      306  CAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 365
Db      3099416  CAGGCGATGACCAACGACCTTCCCTTCCACTTTGATG---GCAGTCCCGCAATTCATC 3099360
Qy      366  GGCCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 425
Db      3099359  GGCCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3099300
Qy      426  AATGATATTTGACGTATGAACTTCAAGATGATGATGATGATGATGATGATGATGAT 485
Db      3099299  AAGGACTTGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3099240
Qy      486  GAAAGCACTGGGACTTGTTCAGAAAGCAAGTGCACGCTGTGATGATGATGATGAT 545
Db      3099239  GAGGCGCTCGGCTGTGCGGGCGGGCGAGGCGCCGAGCTGATGAGACACAGAC 3099180
Qy      546  ACATATGAGAGAAATGAGGTCAATTAATCTGATGATGATGATGATGATGATGATGAT 605
Db      3099179  ACTTACGGGAGCGCTGAGGTGTCAACCATTCGATGATGATGATGATGATGATGATGAT 3099120
Qy      606  CTAGCGCTACAGCTCTTGTCTAGTGTGCAAGATCTCTGCAAGTGAAGAGGAGAAC 665
Db      3099119  CTGGGTGAGAGGGGTTTGGCGAAGTGGCGGAGCTGACCTGCGAGCTCGGGGACCG 3099060
Qy      666  GGAAGAGAGGCAAGTCTCTGTGTCAGAAAGTGTGCTCTGCGCATATTTTANGATT 725
Db      3099059  G--AGCGCGCGAGGTGACAAACGTGACCGCGCGCTGCAACACAAACATCGGCTGG 3099002
Qy      726  GAACTGTGTGTACACTCTAACAAGATG 755
Db      3099001  GCGCGCGGTGTGTCACCGCATACCAACGGG 3098972

```

## RESULT 3

```

US-09-103-840A-2/c

```



```
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, c, g or g
/ US-09-103-840A-2

Query Match      20.1% Score 199.6; DB 3; Length 4403765;
Best Local Similarity 56.8%; Pred. No. 5.1e-46;
Matches 426; Conservative 0; Mismatches 316; Indels 8; Gaps 3;

QY      6  CCAAGTTCCTCTCAGATGTTGGTATGCTGAGAAAAGAACATATGAAATAATGAGACA 65
DB      3094392  CCGGTGGCCCGCGATGTTCCGGGGGGCCGGCCGGGAACATGAAATAATGAGACCC 3094333

QY      66  AAAATTGAACATCTTGCAGAAAATTGATGAGAAAAATCATATAACATTCACTTAATACCCG 125
DB      3094332  ACCGGGAGACATTTCCGCGAAGATCGGCTCAAGAAACCAACACACTCGGTCAACACCCG 3094273

QY      126  TATTCCTCAAGTTCAGATGATATACAGTTAGATGAAGATGAGCATCTTAAGAAAGTTT 185
DB      3094272  TATGACAGCTTCAGAGACAAATACACCTTCGACGACATCTTGCCCTCAAGATGATTTCC 3094213

QY      186  GATTTTTCATCTTACATGTTGTCCCACTTCAGATGCTGTCGACAGCAATTTTG 245
DB      3094212  GACCCGCTGACCAAAATGACAGTCTCTCCACCTCGAGGGGCGGGGCGGTGCTG 3094153

QY      246  GCCAGTGAAGCATTTGTATACAGAAATGAGCTGCAATCCAAAGCTGTGAATTTTGGCA 305
DB      3094152  GCCAGTGAAGATTAAGCTGCGCAACCACTTGGCGGGCGGGGCTGTGAAATCGCGCG 3094093

QY      306  CAAGAAATGATGATGATTTGCAAGCTGTTTGAAGAAAAGCAATTTAAATGTT 365
DB      3094092  CAGGCGATGACCAACGACTTCCTTCACCTTTGATG--GCGATGCCCGCAATATCATC 3094036

QY      366  GCGTTTATATGATTAAGAGAGCTGCAAGAAATGCTATGAGAAATCTGCGCTGACACCA 425
DB      3094035  GCGTACGACATGATGATGCAAGCGGACAAAGGTTTACAGCAATCCGAGACTCGGCCCG 3093976

QY      426  AATGATATGACGTAAATAGAACTTCAAGATTGCTTTTCAACCAAGAACTCTTACTTAT 485
DB      3093975  AAGGACTTCGAGTGTATCAGCTGCGACGACTCTTTCAGCAACAGAGTACTGCTCTAC 3093916

QY      486  GAAAGCATGAGGAGCTGTGTCCAGAGAGCAAGGTGCAAGCTGTTGATAGAGAGTAT 545
DB      3093915  GAGGCGCTGCGCTGTGCGGGCGGGCGGGCGCGCCGAGACTGATGACAGACAAACGAGCC 3093856

QY      546  ACATATGAGAGAAAGTGGTCAATAATCTAGTGTGAGCTGATTTCAAGAGGAGACACCA 605
DB      3093855  ACCTACGGGAGAGCGCTGGGTGCAACCATCGGTTGGGTGATCTCCAAAGGCGCATCCG 3093796

QY      606  CTAGGCGCTACAGAGCTTCTCTCAGTGTGAGAACTTGTCTGCGACGCTGAGAGGAGAGCC 665
DB      3093795  CTGGGTGCGACAGGGGTTGGCGGAGTG--CGAGCTGACCTGCGACGCTGCGCGACCGCC 3093739

QY      666  GAAAAAGAGCAAGTTCCTGTGTGCAAGAGTGAGCTGTGNGCATATTTANGCATTGGAG 725
```

```
DB      3093738  G--AGGCGCGCAGGTGACAAACGACCGCGCGCTGCACAAACATCGGCGCTGGCGG 3093681
QY      726  GAAGTGTGTTGTAACACTCTTACAGATGG 755
DB      3093680  GCGCCGCGGTGTGTCACCGCATACCAACGCG 3093651

RESULT 4
US-08-916-421B-1/c
/ Sequence 1, Application US/08916421B
/ Patent No. 6503729
/ GENERAL INFORMATION:
/ APPLICANT: Bull et al.
/ TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-
/ Patent No. 6503729
/ FILE REFERENCE: jannaschi
/ TITLE OF INVENTION: jannaschi
/ CURRENT APPLICATION NUMBER: US/08/916,421B
/ CURRENT FILING DATE: 1997-08-22
/ PRIOR APPLICATION NUMBER: US 60/024,428
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 1664976
/ TYPE: DNA
/ ORGANISM: Methanococcus jannaschi
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (28252)..(28222)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (28257)..(28258)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (84773)..(84773)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (84808)..(84808)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (84812)..(84812)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (98120)..(98120)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (98159)..(98159)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (98239)..(98239)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (98256)..(98256)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (98343)..(98343)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (103598)..(103598)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (103948)..(103948)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (163385)..(163385)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (191989)..(191989)
/ OTHER INFORMATION: n equals a, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (191995)..(191995)
```

Query	DB	Score	Length	Indels	Gaps
1610105	AAATGGCAAAATATATGAACCAAAAGATGTGATGCTGTAAGTTCAATGCTCTTGTCTA	1610046			
406	AGAAATGTGGCTGACCAAAATGATATTAAGCTTAATAGAACTTCAGATTGCTTTCTA	1610106			
1610165	ACAGCACAAGAGACCATTAACAAGCTTTAAAGCTCTCTAAAGTTGCAAGTGAAGAAAGCATATA	1610106			
346	AAAGCATTTATTAAGATGTTGGCTTTGATATGATGAATGAAGAGCTGCAAGAAATGCTATG	1610166			
1610216	ATATCATCTACATCAACAAAGCAAGGTTCACATCAGATA-----CAATTGCATTAC	1610166			
286	AAAGCTGTGAAATTTTGGCACAAGAAATGATGACTGATTTGCCAAGCTGCTTTGAAGAAA	345			
1610276	GTGTGTGCTGCCCTTATATGTATGTGAAGACAAAAGCTAAAGATTTGTAAATTAAGATG	1610217			
226	GTGTGTGACGACGAATTTTGGCCAGTGAACATTTTGACAGAAATGATGGCTGCATCA	285			
1610336	TCAAATCTTCAACAGTGTCTGAGCTTTAAGATTACTACTACCTCCAGTTTCAGATG	1610277			
1610396	AAAAAGCTTCAAAAAAATGATATGACAAATTCCTTAAGGTTCAATGAGAGAGTTC	1610337			
106	AACATTCAGTTAATTAACCCGTAATTCCTCCAGTTCACAGATGAATACAGTTTGATGAATGA	165			
1610456	ATATGTATGTAGTATGGCTTAACCTTTAGAGAGATTAATCAATGTGAGAGCTTATCATG	1610397			
46	ATATGAAAAATATGGAACAAAAATGGAACATTTGCAAAAAATTTGATGAAAAATCATA	105			
US-08-916-421B-1					
Query Match	9.3%;	Score 92;	DB 4;	Length 1664976;	
Best Local Similarity	49.4%;	Pred. No. 2.6e-15;			
Matches 307;	Conservative 0;	Mismatches 300;	Indels 15;	Gaps 2;	

Oy	466	CAAGCAACTCCTTACTTATGAGGACCTGGGACCTGTCGCAAGAAAGCAACAGTCCAAAGC	525
Db	1610045	TAAATGGTTTAATATTGATGAGGAACTGGTTTCTGTAAAAAGGAGAGGCTGSAAGA	16099866
Oy	526	TGCTTGTAT-----AGAGGAGATATACATATGAGGAAAGTGGGTATTAATCCTAGTG	579
Db	1609985	TAGTTTATGATAAAAAGATACACTATTTGATTATGATGGCTTCCAGCAGTAATCCAAAGTG	16099266
Oy	580	GTGGACTGATTTCAAAGGACACCCACTAGGCGCTACAGTCTTTGCTCAGTGTGCAGAAC	639
Db	1609925	GAGGAGCTGAAAGCTGCTGGAATGATGATTAAGGAGGAGCTACCGCTATTAAGCAGAGTTGGAGGA	16098666
Oy	640	TCGTGTCGCACTGAGAGGGA	661
Db	1609865	TTTACTGGCAGTTAAAGCAGGA	1609944

```

      1
      2
      3
      4
      5
      6
      7
      8
      9
     10
     11
     12
     13
     14
     15
     16
     17
     18
     19
     20
     21
     22
     23
     24
     25
     26
     27
     28
     29
     30
     31
     32
     33
     34
     35
     36
     37
     38
     39
     40
     41
     42
     43
     44
     45
     46
     47
     48
     49
     50
     51
     52
     53
     54
     55
     56
     57
     58
     59
     60
     61
     62
     63
     64
     65
     66
     67
     68
     69
     70
     71
     72
     73
     74
     75
     76
     77
     78
     79
     80
     81
     82
     83
     84
     85
     86
     87
     88
     89
     90
     91
     92
     93
     94
     95
     96
     97
     98
     99
    100
    101
    102
    103
    104
    105
    106
    107
    108
    109
    110
    111
    112
    113
    114
    115
    116
    117
    118
    119
    120
    121
    122
    123
    124
    125
    126
    127
    128
    129
    130
    131
    132
    133
    134
    135
    136
    137
    138
    139
    140
    141
    142
    143
    144
    145
    146
    147
    148
    149
    150
    151
    152
    153
    154
    155
    156
    157
    158
    159
    160
    161
    162
    163
    164
    165
    166
    167
    168
    169
    170
    171
    172
    173
    174
    175
    176
    177
    178
    179
    180
    181
    182
    183
    184
    185
    186
    187
    188
    189
    190
    191
    192
    193
    194
    195
    196
    197
    198
    199
    200
    201
    202
    203
    204
    205
    206
    207
    208
    209
    210
    211
    212
    213
    214
    215
    216
    217
    218
    219
    220
    221
    222
    223
    224
    225
    226
    227
    228
    229
    230
    231
    232
    233
    234
    235
    236
    237
    238
    239
    240
    241
    242
    243
    244
    245
    246
    247
    248
    249
    250
    251
    252
    253
    254
    255
    256
    257
    258
    259
    260
    261
    262
    263
    264
    265
    266
    267
    268
    269
    270
    271
    272
    273
    274
    275
    276
    277
    278
    279
    280
    281
    282
    283
    284
    285
    286
    287
    288
    289
    290
    291
    292
    293
    294
    295
    296
    297
    298
    299
    300
    301
    302
    303
    304
    305
    306
    307
    308
    309
    310
    311
    312
    313
    314
    315
    316
    317
    318
    319
    320
    321
    322
    323
    324
    325
    326
    327
    328
    329
    330
    331
    332
    333
    334
    335
    336
    337
    338
    339
    340
    341
    342
    343
    344
    345
    346
    347
    348
    349
    350
    351
    352
    353
    354
    355
    356
    357
    358
    359
    360
    361
    362
    363
    364
    365
    366
    367
    368
    369
    370
    371
    372
    373
    374
    375
    376
    377
    378
    379
    380
    381
    382
    383
    384
    385
    386
    387
    388
    389
    390
    391
    392
    393
    394
    395
    396
    397
    398
    399
    400
    401
    402
    403
    404
    405
    406
    407
    408
    409
    410
    411
    412
    413
    414
    415
    416
    417
    418
    419
    420
    421
    422
    423
    424
    425
    426
    427
    428
    429
    430
    431
    432
    433
    434
    435
    436
    437
    438
    439
    440
    441
    442
    443
    444
    445
    446
    447
    448
    449
    450
    451
    452
    453
    454
    455
    456
    457
    458
    459
    460
    461
    462
    463
    464
    465
    466
    467
    468
    469
    470
    471
    472
    473
    474
    475
    476
    477
    478
    479
    480
    481
    482
    483
    484
    485
    486
    487
    488
    489
    490
    491
    492
    493
    494
    495
    496
    497
    498
    499
    500
    501
    502
    503
    504
    505
    506
    507
    508
    509
    510
    511
    512
    513
    514
    515
    516
    517
    518
    519
    520
    521
    522
    523
    524
    525
    526
    527
    528
    529
    530
    531
    532
    533
    534
    535
    536
    537
    538
    539
    540
    541
    542
    543
    544
    545
    546
    547
    548
    549
    550
    551
    552
    553
    554
    555
    556
    557
    558
    559
    560
    561
    562
    563
    564
    565
    566
    567
    568
    569
    570
    571
    572
    573
    574
    575
    576
    577
    578
    579
    580
    581
    582
    583
    584
    585
    586
    587
    588
    589
    590
    591
    592
    593
    594
    595
    596
    597
    598
    599
    600
    601
    602
    603
    604
    605
    606
    607
    608
    609
    610
    611
    612
    613
    614
    615
    616
    617
    618
    619
    620
    621
    622
    623
    624
    625
    626
    627
    628
    629
    630
    631
    632
    633
    634
    635
    636
    637
    638
    639
    640
    641
    642
    643
    644
    645
    646
    647
    648
    649
    650
    651
    652
    653
    654
    655
    656
    657
    658
    659
    660
    661
    662
    663
    664
    665
    666
    667
    668
    669
    670
    671
    672
    673
    674
    675
    676
    677
    678
    679
    680
    681
    682
    683
    684
    685
    686
    687
    688
    689
    690
    691
    692
    693
    694
    695
    696
    697
    698
    699
    700
   
```

QY	688	TCGAAGTGCGCTCTGCNGCATAATTANGCATTTGAGAACTGTGTTGAACCTCTA	747
		:  :	
Db	1393	RR	1334
QY	748	CAAGTGGGGTTTCCGGAGCGCGCAGTTCCTTTTGAATCATCAAAATTGAAGCN	807
		:  :	
Db	1333	RR	1274
QY	808	GTTCCAACCAAGCTCTGCAGTNAATNGTTTAAAGNAATAACTGTTTAAAGGCAATT	867
		:  :	
Db	1273	RR	1211
QY	868	GAGAAGAACTNAAGAGGGAAGGGAACAATTGTGAAGAAAAATNCGNGGGA	927
		:  :	
Db	1213	RR	1154
QY	928	TTTTTGCCCTTCAAGGGGAAANAATGCGCCCTGGGGGTTAAAAGANGCCACCTGGGTTG	987
		:  :	
Db	1153	RR	1094
QY	988	GTGGGA  993	
		:  :  :  :	
Db	1093	RRRRR  1088	

US-08-785-048-4  
US-08-785-048-4

1 / Result 6  
2 / Sequence 4, Application US/08785048  
3 / Patent No. 5763246  
4 / GENERAL INFORMATION:  
5 / APPLICANT: Hodgson, John  
6 / APPLICANT: Lawlor, Elizabeth  
7 / TITLE OF INVENTION: No. 5763246el tRNA Synthetase  
8 / NUMBER OF SEQUENCES: 4  
9 / CORRESPONDENCE ADDRESS:  
10 / ADDRESSEE: SmithKline Beecham Corporation  
11 / STREET: 709 Swedeland Road  
12 / CITY: King of Prussia  
13 / STATE: PA  
14 / COUNTRY: USA  
15 / ZIP: 19406-0939  
16 / COMPUTER READABLE FORM:  
17 / MEDIUM TYPE: Diskette  
18 / COMPUTER: IBM Compatible  
19 / OPERATING SYSTEM: DOS  
20 / SOFTWARE: FastSeq for Windows Version 2.0  
21 / CURRENT APPLICATION DATA:  
22 / APPLICATION NUMBER: US/08/785,048  
23 / FILING DATE: 17-JAN-1997  
24 / CLASSIFICATION: 435  
25 / PRIOR APPLICATION DATA:  
26 / APPLICATION NUMBER: 9601099.6  
27 / FILING DATE: 19-JAN-1996  
28 / APPLICATION NUMBER: 9615845.6  
29 / FILING DATE: 27-JUL-1996  
30 / ATTORNEY/AGENT INFORMATION:  
31 / NAME: Gimmil, Edward R  
32 / REGISTRATION NUMBER: 38,891  
33 / REFERENCE/DOCKET NUMBER: P31355-3  
34 / TELECOMMUNICATION INFORMATION:  
35 / TELEPHONE: 610-270-4478  
36 / TELEFAX: 610-270-5090  
37 /  
38 / TELEX:  
39 / INFORMATION FOR SEQ ID NO: 4:  
40 / SEQUENCE CHARACTERISTICS:  
41 / LENGTH: 1644 base pairs  
42 / TYPE: nucleic acid  
43 / STRANDEDNESS: double  
44 / TOPOLOGY: linear  
45 / MOLECULE TYPE: Genomic DNA  
46 / US-08-785-048-4

Query Match 4.0%; Score 39.8; DB 1; Length 1644;  
Best Local Similarity 48.9%; Pred. No. 0.1; Indels 0; Gaps 0;  
Matches 107; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

15 CCTCAGATGTTGGGTATGCTGGAAGAAACATATGAAAAATATGAAACAAAATTTGAA 74  
634 CCTGAAATTAAGATTAATTCGAGAGACCGTTGAAAAGATTGAAAAATTTAGGCGTA 693

QY 75 CACTTGCAAAATTGATGGAATAATCATTAACATTCAGTTATACCCGTAATCCGAG 134  
DB 694 GAATAGGAATGGCTAAATTTGAAAAATGATTTAGCAGAGTTCATGCGATTTTGATAT 753

QY 135 TTCCAAGATGAATACAGTTAGTGAAGATGATGCACTTAAGAGTTTGTATTTTGG 194  
DB 754 TGGTTTAGGAAACATCTTTATATGAAAAAGCGAAATTTCTTGAAGTTTATGCAAAAATG 813

QY 195 ACTATCTTACAATGTTGCCACTTCAGATGCTGTGCA 233  
DB 814 AAGAAATTAGGTTATACGTATGAGCTGATGGCGCTACA 852

RESULT 7  
US-08-996-799-4  
; Sequence 4, Application US/08996799  
; Patent No. 5965416  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 5965416el tRNA Synthetase  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/996,799  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/785,048  
; FILING DATE: 17-JAN-1997  
; APPLICATION NUMBER: 9601099.6  
; FILING DATE: 19-JAN-1996  
; APPLICATION NUMBER: 9615845.6  
; FILING DATE: 27-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimm, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P31355-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1644 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-996-799-4

Query Match 4.0%; Score 39.8; DB 2; Length 1644;  
Best Local Similarity 48.9%; Pred. No. 0.1; Indels 0; Gaps 0;  
Matches 107; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 15 CCTCAGATGTTGGGTATGCTGGAAGAAACATATGAAAAATATGAAACAAAATTTGAA 74  
DB 634 CCTGAAATTAAGATTAATTCGAGAGACCGTTGAAAAGATTGAAAAATTTAGGCGTA 693

QY 75 CACTTGCAAAATTGATGGAATAATCATTAACATTCAGTTATACCCGTAATCCGAG 134  
DB 694 GAATAGGAATGGCTAAATTTGAAAAATGATTTAGCAGAGTTCATGCGATTTTGATAT 753

QY 135 TTCCAAGATGAATACAGTTAGTGAAGATGATGCACTTAAGAGTTTGTATTTTGG 194  
DB 754 TGGTTTAGGAAACATCTTTATATGAAAAAGCGAAATTTCTTGAAGTTTATGCAAAAATG 813

QY 195 ACTATCTTACAATGTTGCCACTTCAGATGCTGTGCA 233  
DB 814 AAGAAATTAGGTTATACGTATGAGCTGATGGCGCTACA 852

RESULT 8  
US-08-785-048-1  
; Sequence 1, Application US/08785048  
; Patent No. 5763246  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 5763246el tRNA Synthetase  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,048  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9601099.6  
; FILING DATE: 19-JAN-1996  
; APPLICATION NUMBER: 9615845.6  
; FILING DATE: 27-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimm, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P31355-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1662 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-785-048-1

Query Match 4.0%; Score 39.8; DB 1; Length 1662;  
Best Local Similarity 48.9%; Pred. No. 0.1; Indels 0; Gaps 0;  
Matches 107; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 15 CCTCAGATGTTGGGTATGCTGGAAGAAACATATGAAAAATATGAAACAAAATTTGAA 74  
DB 652 CCTGAAATTAAGATTAATTCGAGAGACCGTTGAAAAGATTGAAAAATTTAGGCGTA 711



```
Query Match      3 8%; Score 38.2; DB 4; Length 1368;
Best Local Similarity 56.9%; Pred.No.0.27; Mismatches 53; Indels 0; Gaps 0;
Matches 70; Conservative 0;

QY      387 GCTGCAGAGAAATGCTATGAGAATCTGCCTGACACCAATGATATGACGTAATAGAA 446
        |||||
DB      1072 GCTGCACGCGAAGCCCTTGAGAAAAGCTGTTTAACATGATGATATGACCTATTGAA 1131
        |||||

QY      447 CTTACAGATTGCTTTTCTTCAACAGACCTCTTACTTATGAAAGCACTGAGACTGTGCA 506
        |||||
DB      1132 GTTAAAGCAAGCATTTGACCTGCTGATGCGTTTATTAATGAACTTATGATGCTGCT 1191
        |||||

QY      507 GAA 509
        |||
DB      1192 GAA 1194

RESULT 12
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Buit et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
```

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163585)..(163585)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191889)..(191889)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
```

```

NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1345473)..(1345473)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, c, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, c, c, or g
US-08-916-421B-1

```

```

Query Match 3.8%; Score 37.8; DB 4; Length 1664576;
Best Local Similarity 52.2%; Pred. No. 9.7;
Matches 84; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 43 AACATATGGAATAATATGGAACAAATGACACTTTCGCAAAATTTGGATGGAATAATC 102
DB 732922 AAGATTGTTGATATATATATATGATTTAAACCTTGAATAGTTAAAGTTAAAGG 732981

QY 103 ATAAACATTACGTTAATAACCCGTAATTCACAGTCCAGATGAATACAGTTAGATGAAG 162
DB 732982 AAAATGAATGAAATTAAGCTAAATCTCTTAAATCTGAAATATATATGATTCAT 733041

QY 163 TGATGCATCTNAAGAGAGTTTGTGATTTTGTACTATCTTA 203
DB 733042 TAGTGTAGCTCCAGAGATGAAGACATTTTATATATTTA 733082

```

```

RESULT 13
US-09-252-991A-13040
; Sequence 13040, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```

```

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13040
TYPE: DNA
LENGTH: 1353
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13040

```

```

Query Match 3.8%; Score 37.4; DB 4; Length 1353;
Best Local Similarity 49.2%; Pred. No. 0.45;
Matches 98; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```

```

QY 406 AGAATCTGCGCTGACACCAATGATATGACGTATGAACTTACGATTCCTTTCTA 465
DB 1022 AGATGCCGGGTGTGGGACGCCGACATCGACTGTGATGCTCTACGACGCTTTACCA 1081

QY 466 CCAACGAACCTCTTACTATATGAGCAGCTGGACTGTCCAGAGACAAAGTGCAAGC 525
DB 1082 TCACACCTCTGTTCTTCTGAGACCTGGGATTTGGCCAAAGGCGGCGCCCTT 1141

QY 526 TGGTATAGAGAGATATATCATATGAGAGAAAGTGATATAATCTAGTGTGAC 585
DB 1142 TCGTCAGAGGGGAGACGATCGCGCGCGGAGTTGGCGGTGAACCAAGGTGGG 1201

QY 586 TGATTTCAAAGGACACCC 604
DB 1202 GCGTGTCTGCGTGCATCC 1220

```

```

RESULT 14
US-09-252-991A-12438/c
; Sequence 12438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12438
LENGTH: 1410
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: unsure
LOCATION: (93)..(101)
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-12438

```

```

Query Match 3.8%; Score 37.4; DB 4; Length 1410;
Best Local Similarity 49.2%; Pred. No. 0.46;
Matches 98; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```

```

QY 406 AGAATCTGCGCTGACACCAATGATATGACGTATGAACTTACGATTCCTTTCTA 465
DB 461 AGATGCCGGGTGTGGGACGCCGACATCGACTGTGATGCTCTACGACGCTTTACCA 402

QY 466 CCAACGAACCTCTTACTATATGAGCAGCTGGACTGTCCAGAGACAAAGTGCAAGC 525

```

Db 401 TCACACCCCTGCTTCTCTGAGACCTGGATTTCGCCCAAGGGCGAGGCCCAT 342  
QY 526 TGGTTATATGAGGATTAATACATATGAGGAAAGTGGTCATAATCTAGTGTGAC 585  
Db 341 TCGTCCAGGGCGGACCATCGCCCGGGCGAGTTGGCGGTGAACACCAACGCTGGCG 282  
QY 586 TGAATTCACAAAGGACACCC 604  
Db 281 GGCTGTCTGCTGTCATCC 263

## RESULT 15

US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-Zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelisen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 3.6%; Score 35.8; DB 2; Length 19124;  
Best Local Similarity 52.3%; Pred. No. 4.5;  
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 52 AAAAAATATGAAACAAAAATGAAACACTTGCACAAAAATGGATGAGAAAAATCATTAACATT 111  
Db 3418 AAAAAATGAGAAAAAATGCAATATATAAATTAAGGAATAGATATACGATGTC 3477  
QY 112 CAGTTATATACCCGATTTCCAGTTCACAGATGATACAGTTAGATGAAGTATGCGCAT 171  
Db 3478 GAGCTATTTAGCGGTAATTTAAAGTATGGAATTTTCAATTAATGCTATGATCATTT 3537  
QY 172 CTAAGAGATTTTGAATTTTGTACTATCTT 202

Db 3538 GATAATTAATTTTTTTTATATAATTAATTAATTT 3568

Search completed: November 27, 2003, 12:32:29  
Job time: 85.3681 secs





FEATURES  
source 1. .843 Location/Qualifiers  
BASE COUNT 237 a 184 c 89 g 215 t 118 others  
ORIGIN

Query Match 86.0%; Score 725; DB 6; Length 843;  
Best Local Similarity 100.0%; Pred. No. 7.6e-139;  
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCAAAAAAATTTTAAATTCCTATTAACCTCCGCAAAACATTAATTACCC 60  
DB 1 GGGCAAAAAAATTTTAAATTCCTATTAACCTCCGCAAAACATTAATTACCC 60  
QY 61 TATNNCNCNGANTTTNANAANTACCTTNNNTTAAAAAACCTNGAAAAAATA 120  
DB 61 TATNNCNCNGANTTTNANAANTACCTTNNNTTAAAAAACCTNGAAAAAATA 120  
QY 121 TNGCAATTAATTAACCTTNGAAANGAAATTTTACGAANGACGAANCTNT 180  
DB 121 TNGCAATTAATTAACCTTNGAAANGAAATTTTACGAANGACGAANCTNT 180  
QY 181 AATTGAANTAAATTAATTTNGAANGGCGNCGAACCACCTTATGTCAT 240  
DB 181 AATTGAANTAAATTAATTTNGAANGGCGNCGAACCACCTTATGTCAT 240  
QY 241 ATCTTAAANGAGGNNTTNANNACTAATNCCNGATTTTCAATPANGAANCCNN 300  
DB 241 ATCTTAAANGAGGNNTTNANNACTAATNCCNGATTTTCAATPANGAANCCNN 300  
QY 301 AAAANTTTTAAATTTTAAATAACCGGNTTCAACCCCGATCANATTCCTT 360  
DB 301 AAAANTTTTAAATTTTAAATAACCGGNTTCAACCCCGATCANATTCCTT 360  
QY 361 TGGATTGGGAAAAAATNCNGTCCNNATACCNNGAANGCAANTTTTAAAT 420  
DB 361 TGGATTGGGAAAAAATNCNGTCCNNATACCNNGAANGCAANTTTTAAAT 420  
QY 421 ACCCCCTTAAATTTTAAANGCTATNGAANGAANGAANGAANGAANGAANG 480  
DB 421 ACCCCCTTAAATTTTAAANGCTATNGAANGAANGAANGAANGAANGAANG 480  
QY 481 TTNNCGGCGACGNGGCGNTGNTTCTTACTTANCCCGCAAGGAANNCCT 540  
DB 481 TTNNCGGCGACGNGGCGNTGNTTCTTACTTANCCCGCAAGGAANNCCT 540  
QY 541 CNGAANGCTCCNCAAAATACCTTAANTATCTTGTACCAACCAACCACT 600  
DB 541 CNGAANGCTCCNCAAAATACCTTAANTATCTTGTACCAACCAACCACT 600  
QY 601 TTTACNTANTCTTGGGATTTAAGGGTCCCAATTTNATCCGAACTTTCC 660  
DB 601 TTTACNTANTCTTGGGATTTAAGGGTCCCAATTTNATCCGAACTTTCC 660  
QY 661 NAACATANTTAACATTTACCTGTAGGCGACGNTGTCANTNCGGAANCA 720  
DB 661 NAACATANTTAACATTTACCTGTAGGCGACGNTGTCANTNCGGAANCA 720  
QY 721 ANTNTTCCCGGCGCTTCCCGGACCTTGGGAAAAAGGAGATNGTCC 780  
DB 721 ANTNTTCCCGGCGCTTCCCGGACCTTGGGAAAAAGGAGATNGTCC 780  
QY 781 AAAACAACCTTCCCGGCGCTTGGGCGAGGNTTNTCCGCTTAAATCGA 840  
DB 781 AAAACAACCTTCCCGGCGCTTGGGCGAGGNTTNTCCGCTTAAATCGA 840  
QY 841 AAG 843  
DB 841 AAG 843

LOCUS BD079831 843 bp DNA linear PAT 27-AUG-2002  
DEFINITION Cancer-associated nucleic acids and polypeptides.  
ACCESSION BD079831  
VERSION BD079831.1 GI:22625434  
KEYWORDS JP 2001516009-A/497.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Old,L.J., Scanlan,M.J., Stockert,E., Gure,A., Chen,Y.T., Gout,I.,  
Oghare,M., Obara,Y., Preunsdorh,M., Tureci,O. and Sahin,U.  
TITLE Cancer-associated nucleic acids and polypeptides  
JOURNAL Patent: JP 2001516009-A 497 25-SEP-2001;  
LUDWIG INSTITUTE FOR CANCER RESEARCH  
COMMENT  
OS Homo sapiens (human)  
PN JP 2001516009-A/497  
PD 25-SEP-2001  
PR 15-JUL-1998 JP 2000503425  
PR 17-JUL-1997 US 08/896164,10-OCT-1997 US 60/061599 PR  
10-OCT-1997 US 60/061765,10-OCT-1997 US 08/948705 PR  
11-OCT-1997 GB 9721697.2,22-JUN-1998 US 09/102322 PI  
J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI  
CHEN, YUAN GOUT, MICHAEL O'HARE, YUICHI OBARA, MICHAEL PREUNSDORH, PI  
PI OZLEM TURECI,  
PI UGUR SAHIN,  
PC G01N33/574,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61K48/00, PC  
A61P35/00,  
PC C07K14/82,C07K16/32,C12N15/09//C07K16/46,C12P21/08,A61K37/02,  
PC C12N15/00  
CC Cancer-associated nucleic acids and polypeptides. FH Key  
Location/Qualifiers  
FT source 1. .843 /organism='Homo sapiens (human)'.  
FEATURES  
source 1. .843 Location/Qualifiers  
BASE COUNT 237 a 184 c 89 g 215 t 118 others  
ORIGIN

Query Match 86.0%; Score 725; DB 6; Length 843;  
Best Local Similarity 100.0%; Pred. No. 7.6e-139;  
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCAAAAAAATTTTAAATTCCTATTAACCTCCGCAAAACATTAATTACCC 60  
DB 1 GGGCAAAAAAATTTTAAATTCCTATTAACCTCCGCAAAACATTAATTACCC 60  
QY 61 TATNNCNCNGANTTTNANAANTACCTTNNNTTAAAAAACCTNGAAAAAATA 120  
DB 61 TATNNCNCNGANTTTNANAANTACCTTNNNTTAAAAAACCTNGAAAAAATA 120  
QY 121 TNGCAATTAATTAACCTTNGAAANGAAATTTTACGAANGACGAANCTNT 180  
DB 121 TNGCAATTAATTAACCTTNGAAANGAAATTTTACGAANGACGAANCTNT 180  
QY 181 AATTGAANTAAATTAATTTNGAANGGCGNCGAACCACCTTATGTCAT 240  
DB 181 AATTGAANTAAATTAATTTNGAANGGCGNCGAACCACCTTATGTCAT 240  
QY 241 ATCTTAAANGAGGNNTTNANNACTAATNCCNGATTTTCAATPANGAANCCNN 300  
DB 241 ATCTTAAANGAGGNNTTNANNACTAATNCCNGATTTTCAATPANGAANCCNN 300  
QY 301 AAAANTTTTAAATTTTAAATAACCGGNTTCAACCCCGATCANATTCCTT 360  
DB 301 AAAANTTTTAAATTTTAAATAACCGGNTTCAACCCCGATCANATTCCTT 360  
QY 361 TGGATTGGGAAAAAATNCNGTCCNNATACCNNGAANGCAANTTTTAAAT 420

|||||  
Db TGGATTGGGAAAAAATNCNGTTCNNATACNNNGAANNCAAAATTTTAAATTTTAA 420  
Oy 421 ACCCCCTTAATTTTAAANCTATNGAAAAANTNGATTANNAGCTTAATTTGCCAACCTTAN 480  
Db 421 ACCCCCTTAATTTTAAANCTATNGAAAAANTNGATTANNAGCTTAATTTGCCAACCTTAN 480  
Oy 481 TTTCNGGCGACCGNGGCGTNTGCTTACTTANTCCCGGCAAGGAANNCCCTTAN 540  
Db 481 TTTCNGGCGACCGNGGCGTNTGCTTACTTANTCCCGGCAAGGAANNCCCTTAN 540  
Oy 541 CNGAANCTCCNCAAAATTAACCTTAATATCTTGTATACCAANCAAAACCTTTTNG 600  
Db 541 CNGAANCTCCNCAAAATTAACCTTAATATCTTGTATACCAANCAAAACCTTTTNG 600  
Oy 601 TTATCTANTATCTTGGGATTTAAGCGGTCCCAATTTNAATCCNCAACCAATTTTCCCC 660  
Db 601 TTATCTANTATCTTGGGATTTAAGCGGTCCCAATTTNAATCCNCAACCAATTTTCCCC 660  
Oy 661 NAACCATANTTAACATTTTAACTTGTATAGGCGNCGTNGTTGCANTCCGCAANCACT 720  
Db 661 NAACCATANTTAACATTTTAACTTGTATAGGCGNCGTNGTTGCANTCCGCAANCACT 720  
Oy 721 ANTNTTCCCGGCGCTTTCCTCCCGGCAAGGATNGTCCCGCTTAA 780  
Db 721 ANTNTTCCCGGCGCTTTCCTCCCGGCAAGGATNGTCCCGCTTAA 780  
Oy 781 AAAACACCTTCCCGGCTTTCCTCCCGGCAAGGATNGTCCCGCTTAA 840  
Db 781 AAAACACCTTCCCGGCTTTCCTCCCGGCAAGGATNGTCCCGCTTAA 840  
Oy 841 AAG 843  
Db 841 AAG 843

RESULT 3  
HSCPT16/c  
LOCUS HSCPT16 1052 bp DNA linear PRI 25-JAN-1997  
DEFINITION Human sterol carrier protein-X/sterol carrier protein-2  
ACCESSION U11313  
VERSION U11313.1 GI:532077  
KEYWORDS  
SEGMENT 1 of 16  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1052) Chiba, T., Renner, H., Pfeiffer, S.M., He, Z., Yamamoto, R., Holt, J.A., Billheimer, J.T. and Strauss, J.F. III. The structure of the human sterol carrier protein X/sterol carrier protein 2 gene (SCP2) Genomics 24 (2), 370-374 (1994)  
TITLE JOURNAL MEDLINE PUBMED  
PUBMED 7698762  
REFERENCE 2 (bases 1 to 1052) Strauss, J.F. III. Direct Submision  
AUTHORS Submitted (24-JUN-1994) Jerome F. Strauss III, Department of Obstetrics and Gynecology, Division of Reproductive Biology, University of Pennsylvania Medical Center, 778 Clinical Research Building, 422 Curie Boulevard, Philadelphia, PA 19104-6142, USA  
JOURNAL TITLE  
FEATURES  
SOURCE 1..1052  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="Caxon:9606"  
/chromosome="1"  
/map="1p32"  
/cfeature\_type="1iver"  
join(U11297.1:1..3666,U11299.1:1..108,U11300.1:1..122,

U11301.1:1..182,U11302.1:1..115,U11303.1:1..177,  
U11304.1:1..114,U11305.1:1..137,U11306.1:1..201,  
U11307.1:1..198,U11308.1:1..158,U11309.1:1..1725,  
U11310.1:1..153,U11311.1:1..180,U11312.1:1..130,1..1034)  
/partial  
/gene="SCP-X/SCP-2"  
join(U11297.1:3452..3641,U11299.1:26..83,U11300.1:26..97,  
U11301.1:26..157,U11302.1:26..90,U11303.1:26..152,  
U11304.1:26..89,U11305.1:26..112,U11306.1:26..176,  
U11307.1:26..173,U11308.1:26..133,U11309.1:1547..1700,  
U11310.1:26..128,U11311.1:26..155,U11312.1:26..105,  
26..1034)  
/gene="SCP-X/SCP-2"  
/product="sterol carrier protein-X/sterol carrier  
protein-2"  
join(U11297.1:3573..3641,U11299.1:26..83,U11300.1:26..97,  
U11301.1:26..157,U11302.1:26..90,U11303.1:26..152,  
U11304.1:26..89,U11305.1:26..112,U11306.1:26..176,  
U11307.1:26..173,U11308.1:26..133,U11309.1:1547..1700,  
U11310.1:26..128,U11311.1:26..155,U11312.1:26..105,  
26..121)  
/gene="SCP-X/SCP-2"  
/codon\_start=1  
/product="sterol carrier protein-X/sterol carrier  
protein-2"  
/protein\_id="AAB41286.1"  
/db\_xref="GI:1773239"  
/translation="MSSSPWEPALRRVYVVGNTKFKRGAENSRDYPDLAEAGK  
KALADAIQGVSAVDAOCVGVFGDSTCGPRAIYHSLGTGPIPIINNNCAVSTALP  
MAROLIQGVSAVDAOCVGVFGDSTCGPRAIYHSLGTGPIPIINNNCAVSTALP  
OMFGVAKHEHEKATKIEHFAKIGMKHKSIVNVPQFODEYSLDVMSKEVDF  
LTIQCCPTSDGAAATIIASAFAPOKVGLOKAVEIIIAOEMTDLPSPFEKSTIKM  
GPDMSKNAKRCYKSGITPNDIVIEIHDFTREBELTYELGICPEGOGATTVDRG  
DNIVGKVINPSPGLISKHPIDATGALQCAELCWQJRGAKRQVGAQVLAQHL  
DIGAVVVTLYMGFPEBAASRTHQIEAVPSSSDPFGKALVKELEKLEEGEO  
FVKKIGIFAFKVDGPGKEATVVDVQNGKSVLPSPSKADCTTITMADSDFLAM  
TGRNPOSAPFQGLKITGNMGLAMKIQNLQGNALTL"  
26..1034  
/gene="SCP-X/SCP-2"  
/note="nucleotides 122-1027 are not translated"  
/number=16

exon  
BASE COUNT 352 a 157 c 164 g 379 t  
ORIGIN  
Query Match 59.9%; Score 505; DB 9; Length 1052;  
Best Local Similarity 71.8%; Pred. No. 1.1e-93;  
Matches 610; Conservative 0; Mismatches 233; Indels 6; Gaps 3;  
Oy 1 GCCCAAAAANTTAATTTCTTAAACCTTCTCCGCAANCAATTAATTNACC 60  
Db 1028 GCCCAAAAAGTTATTTTAACTTTCTTCTCCGCAANCAATTAATTATCC 969  
Oy 61 TATNNCNCNGANTTANAAANTACCTTNNNTTAAAAAACCCTNGAAAAAATTA 120  
Db 968 TATATCTACGAAATTTAAGAAATACATTAAGTTTGAABAACTAGAAABAAAGATA 909  
Oy 121 TNGCAATANTTAACCTTCTTGAANGAATTTTACCAANGACNGAANCNTTNT 180  
Db 908 ATGCAGATAATTAACCTTACATGAAAAAGAAAAATTAATACAAAGACTGAGAACCTTAT 849  
Oy 181 AATTNGAANTTAATTAANTTNGAANGCGCNCNGAANGAACCANCTTNTAGTCCATTT 240  
Db 848 AAATGGAATAGATTAATTAATTTGAAAACCTGATCTGAAAGCAAACTTATTTGTCATTT 789  
Oy 241 ATCTNAANGAGGNNTTTANNACTAATNCCNGATTTTCCATANGAANCCNNNTT 300  
Db 788 ATTCTATAGATGGGTTTATGACTAATACATGATTTTTCATTAAGAAACCATGTT 729  
Oy 301 AAAANTTTTNAATTTAAAAATAACCCGNTTCCACCCCGATCANATTCCTTNAAT 360  
Db 728 AAAATATTTTATTTTAAAAATAAGCCTGTGTCACACCTGATCATATTTCTTTATTT 669  
Oy 361 TGGATTGGGAAAAAATNCNGTTCNNATACNNNGAANNCAAAATTTTAAATTTTAA 420



Oy	1.6	AATTNGAANTAAATTTATNTTNGAANGGCGNCCGAAACGAACCTTNAAGTCCAAAT	240
Db	47460	AAATTGAATGAGATTATAATTGGAAACAGCATCTGAAAGCAACTTATATGTTCAAT	475119
Oy	241	ATCTTNAANGGGGNNNTTANNACTAATCCNGATTTTCCAAATNANGAANCCNNNT	300
Db	47520	ATTCTTAATGATGCGTGTATTATGACTAATACATGATTTTCAATAGGAAACCCATGTT	475797
Oy	301	AAAAATNTTNTATTTTAAAAATACCCGNGTCCAAACCCGATCAATTCCTTNAAT	360
Db	47580	AAAAATTTTATTTTAAAAATAGCCTGTGTTCAAGCTCTGATCATATTTCTTTAT	476339
Oy	361	TGGATTGGGGAABAAAATNCGTCCNNATACCCNNGAANNCAATTTTAAATTTTA	420
Db	47640	TTGATTTGGGAAGAAATACGTGTTTCGAATAGCATGAATATGCAAAATTTTAAATTTTA	476997
Oy	421	ACCCCTCTANTTTTAAAAACATANGAATAATNATANNACTTGAATTCG-CAACCTTA	479
Db	47700	ATCTCACTAATTTTAAABAACATATTGAGAAATTTGATTAATGACATGAAGTGCACAACTTA	479555
Oy	480	NTTNGCGGCCACGCGTGGCGNTNGTNTCTTACTTAATTCCTCCCAAGGAANNCTTAA	539
Db	47760	ATTACTGCGCAGCTGTGGCATGTGTGTTCTTACTTAAGTTCTCCCAAGGAAACCTTAA	478119
Oy	540	NCNGAANCCTCCNCCAAATAACCTTAAATCTTGTGTAAACAAANCAAAACCTTTTN	599
Db	47820	ACTGAATCTTCAGAGAAATATCTTAAATATCTTGTGAAGAAACAAAGCTTTT	478797
Oy	600	GTTTACATTAATCTTTGGGATTTAACGGAGTCCCAATTNATCCGAACCCATTTTCCCC	659
Db	47880	GTTTACATAGTCTTTGAGATTTTACTGTCTTAATTTTATCTGAAACCTCAATTTTACC	479339
Oy	660	CNAACATANTTACCATTTTAACTTGGTAAAGCCAGTNGTTGCANTCCGCAANCG	719
Db	47940	CAGACCAATTAATACATATTAACTTTGTAAATGACAGTTGTATGCAATTCGCAAGCAG	479939
Oy	720	TANTNTTCC--CCNGGNCCTTTCCCCGAANCCTTTGGAAAAACGGATNGTCCCCCT	777
Db	48000	TAGATATCAATCAGGCTCTATTAACCCAGAGCAATAGAAAAACAGATTTGTGTACACCA	480539
Oy	778	T---AAAAAACCACTTCCGCCNCTTTGGCCGAGNNNTTNTCCGCTTAATTCGAAC	834
Db	48060	TTAAGAAACAATTAATCCACGCTATTTGCCAATGTGTAGTTTCAGTCTAAATTTCTGAC	481119
Oy	835	AATTAAG 843	
Db	48120	AATTAATG 48128	
RESULT 5			
AC022728/c			
LOCUS	175046 bp	DNA	linear
DEFINITION	Homo sapiens chromosome 1 clone RP11-310J14 map 1, WORKING DRAFT	HTG 24-AUG-2002	
SEQUENCE	34 unordered pieces.		
AC022728			
AC022728.4	GI:7249198		
HTG: HTGS PHASE1; HTGS _DRAFT.			
KEYWORDS	Homo sapiens (human)		
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
1	(bases 1 to 175046)		
2	(bases 1 to 175046)		
3	(bases 1 to 175046)		
4	(bases 1 to 175046)		
5	(bases 1 to 175046)		
6	(bases 1 to 175046)		
7	(bases 1 to 175046)		
8	(bases 1 to 175046)		
9	(bases 1 to 175046)		
10	(bases 1 to 175046)		
11	(bases 1 to 175046)		
12	(bases 1 to 175046)		
13	(bases 1 to 175046)		
14	(bases 1 to 175046)		
15	(bases 1 to 175046)		
16	(bases 1 to 175046)		
17	(bases 1 to 175046)		
18	(bases 1 to 175046)		
19	(bases 1 to 175046)		
20	(bases 1 to 175046)		
21	(bases 1 to 175046)		
22	(bases 1 to 175046)		
23	(bases 1 to 175046)		
24	(bases 1 to 175046)		
25	(bases 1 to 175046)		
26	(bases 1 to 175046)		
27	(bases 1 to 175046)		
28	(bases 1 to 175046)		
29	(bases 1 to 175046)		
30	(bases 1 to 175046)		
31	(bases 1 to 175046)		
32	(bases 1 to 175046)		
33	(bases 1 to 175046)		
34	(bases 1 to 175046)		
35	(bases 1 to 175046)		
36	(bases 1 to 175046)		
37	(bases 1 to 175046)		
38	(bases 1 to 175046)		
39	(bases 1 to 175046)		
40	(bases 1 to 175046)		
41	(bases 1 to 175046)		
42	(bases 1 to 175046)		
43	(bases 1 to 175046)		
44	(bases 1 to 175046)		
45	(bases 1 to 175046)		
46	(bases 1 to 175046)		
47	(bases 1 to 175046)		
48	(bases 1 to 175046)		
49	(bases 1 to 175046)		
50	(bases 1 to 175046)		
51	(bases 1 to 175046)		
52	(bases 1 to 175046)		
53	(bases 1 to 175046)		
54	(bases 1 to 175046)		
55	(bases 1 to 175046)		
56	(bases 1 to 175046)		
57	(bases 1 to 175046)		
58	(bases 1 to 175046)		
59	(bases 1 to 175046)		
60	(bases 1 to 175046)		
61	(bases 1 to 175046)		
62	(bases 1 to 175046)		
63	(bases 1 to 175046)		
64	(bases 1 to 175046)		
65	(bases 1 to 175046)		
66	(bases 1 to 175046)		
67	(bases 1 to 175046)		
68	(bases 1 to 175046)		
69	(bases 1 to 175046)		
70	(bases 1 to 175046)		
71	(bases 1 to 175046)		
72	(bases 1 to 175046)		
73	(bases 1 to 175046)		
74	(bases 1 to 175046)		
75	(bases 1 to 175046)		
76	(bases 1 to 175046)		
77	(bases 1 to 175046)		
78	(bases 1 to 175046)		
79	(bases 1 to 175046)		
80	(bases 1 to 175046)		
81	(bases 1 to 175046)		
82	(bases 1 to 175046)		
83	(bases 1 to 175046)		
84	(bases 1 to 175046)		
85	(bases 1 to 175046)		
86	(bases 1 to 175046)		
87	(bases 1 to 175046)		
88	(bases 1 to 175046)		
89	(bases 1 to 175046)		
90	(bases 1 to 175046)		
91	(bases 1 to 175046)		
92	(bases 1 to 175046)		
93	(bases 1 to 175046)		
94	(bases 1 to 175046)		
95	(bases 1 to 175046)		
96	(bases 1 to 175046)		
97	(bases 1 to 175046)		
98	(bases 1 to 175046)		
99	(bases 1 to 175046)		
100	(bases 1 to 175046)		
101	(bases 1 to 175046)		
102	(bases 1 to 175046)		
103	(bases 1 to 175046)		
104	(bases 1 to 175046)		
105	(bases 1 to 175046)		
106	(bases 1 to 175046)		
107	(bases 1 to 175046)		
108	(bases 1 to 175046)		
109	(bases 1 to 175046)		
110	(bases 1 to 175046)		
111	(bases 1 to 175046)		
112	(bases 1 to 175046)		
113	(bases 1 to 175046)		
114	(bases 1 to 175046)		
115	(bases 1 to 175046)		
116	(bases 1 to 175046)		
117	(bases 1 to 175046)		
118	(bases 1 to 175046)		
119	(bases 1 to 175046)		
120	(bases 1 to 175046)		
121	(bases 1 to 175046)		
122	(bases 1 to 175046)		
123	(bases 1 to 175046)		
124	(bases 1 to 175046)		
125	(bases 1 to 175046)		
126	(bases 1 to 175046)		
127	(bases 1 to 175046)		
128	(bases 1 to 175046)		
129	(bases 1 to 175046)		
130	(bases 1 to 175046)		
131	(bases 1 to 175046)		
132	(bases 1 to 175046)		
133	(bases 1 to 175046)		
134	(bases 1 to 175046)		
135	(bases 1 to 175046)		
136	(bases 1 to 175046)		
137	(bases 1 to 175046)		
138	(bases 1 to 175046)		
139	(bases 1 to 175046)		
140	(bases 1 to 175046)		
141	(bases 1 to 175046)		
142	(bases 1 to 175046)		
143	(bases 1 to 175046)		
144	(bases 1 to 175046)		
145	(bases 1 to 175046)		
146	(bases 1 to 175046)		
147	(bases 1 to 175046)		
148	(bases 1 to 175046)		
149	(bases 1 to 175046)		
150	(bases 1 to 175046)		
151	(bases 1 to 175046)		
152	(bases 1 to 175046)		
153	(bases 1 to 175046)		
154	(bases 1 to 175046)		
155	(bases 1 to 175046)		
156	(bases 1 to 175046)		
157	(bases 1 to 175046)		
158	(bases 1 to 175046)		
159	(bases 1 to 175046)		
160	(bases 1 to 175046)		
161	(bases 1 to 175046)		
162	(bases 1 to 175046)		
163	(bases 1 to 175046)		
164	(bases 1 to 175046)		
165	(bases 1 to 175046)		
166	(bases 1 to 175046)		
167	(bases 1 to 175046)		
168	(bases 1 to 175046)		
169	(bases 1 to 175046)		
170	(bases 1 to 175046)		
171	(bases 1 to 175046)		
172	(bases 1 to 175046)		
173	(bases 1 to 175046)		
174	(bases 1 to 175046)		
175	(bases 1 to 175046)		
176	(bases 1 to 175046)		
177	(bases 1 to 175046)		
178	(bases 1 to 175046)		
179	(bases 1 to 175046)		
180	(bases 1 to 175046)		
181	(bases 1 to 175046)		
182	(bases 1 to 175046)		
183	(bases 1 to 175046)		
184	(bases 1 to 175046)		
185	(bases 1 to 175046)		
186	(bases 1 to 175046)		
187	(bases 1 to 175046)		
188	(bases 1 to 175046)		
189	(bases 1 to 175046)		
190	(bases 1 to 175046)		
191	(bases 1 to 175046)		
192	(bases 1 to 175046)		
193	(bases 1 to 175046)		
194	(bases 1 to 175046)		
195	(bases 1 to 175046)		
196	(bases 1 to 175046)		
197	(bases 1 to 175046)		
198	(bases 1 to 175046)		
199	(bases 1 to 175046)		
200	(bases 1 to 175046)		
201	(bases 1 to 175046)		
202	(bases 1 to 175046)		
203	(bases 1 to 175046)		
204	(bases 1 to 175046)		
205	(bases 1 to 175046)		
206	(bases 1 to 175046)		
207	(bases 1 to 175046)		
208	(bases 1 to 175046)		
209	(bases 1 to 175046)		
210	(bases 1 to 175046)		
211	(bases 1 to 175046)		
212	(bases 1 to 175046)		
213	(bases 1 to 175046)		
214	(bases 1 to 175046)		
215	(bases 1 to 175046)		
216	(bases 1 to 175046)		
217	(bases 1 to 175046)		
218	(bases 1 to 175046)		
219	(bases 1 to 175046)		
220	(bases 1 to 175046)		
221	(bases 1 to 175046)		
222	(bases 1 to 175046)		
223	(bases 1 to 175046)		
224	(bases 1 to 175046)		
225	(bases 1 to 175046)		
226	(bases 1 to 175046)		
227	(bases 1 to 175046)		
228	(bases 1 to 175046)		
229	(bases 1 to 175046)		
230	(bases 1 to 175046)		
231	(bases 1 to 175046)		
232	(bases 1 to 175046)		
233	(bases 1 to 175046)		
234	(bases 1 to 175046)		
235	(bases 1 to 175046)		
236	(bases 1 to 175046)		
237	(bases 1 to 175046)		
238	(bases 1 to 175046)		
239	(bases 1 to 175046)		
240	(bases 1 to 175046)		
241	(bases 1 to 175046)		
242	(bases 1 to 175046)		
243	(bases 1 to 175046)		
244	(bases 1 to 175046)		
245	(bases 1 to 175046)		
246	(bases 1 to 175046)		
247	(bases 1 to 175046)		
248	(bases 1 to 175046)		
249	(bases 1 to 175046)		
250	(bases 1 to 175046)		
251	(bases 1 to 175046)		
252	(bases 1 to 175046)		
253	(bases 1 to 175046)		
254	(bases 1 to 175046)		
255	(bases 1 to 175046)		
256	(bases 1 to 175046)		
257	(bases 1 to 175046)		
258	(bases 1 to 175046)		
259	(bases 1 to 175046)		
260	(bases 1 to 175046)		
261	(bases 1 to 175046)		
262	(bases 1 to 175046)		
263	(bases 1 to 175046)		
264	(bases 1 to 175046)		
265	(bases 1 to 175046)		
266	(bases 1 to 175046)		
267	(bases 1 to 175046)		
268	(bases 1 to 175046)		
269	(bases 1 to 175046)		
270	(bases 1 to 175046)		
271	(bases 1 to 175046)		
272	(bases 1 to 175046)		
273	(bases 1 to 175046)		
274	(bases 1 to 175046)		
275	(bases 1 to 175046)		
276	(bases 1 to 175046)		
277	(bases 1 to 175046)		
278	(bases 1 to 175046)		
279	(bases 1 to 175046)		
280	(bases 1 to 175046)		
281	(bases 1 to 175046)		
282	(bases 1 to 175046)		
283	(bases 1 to 175046)		
284	(bases 1 to 175046)		
285	(bases 1 to 175046)		
286	(bases 1 to 175046)		
287	(bases 1 to 175046)		
288	(bases 1 to 175046)		
289	(bases 1 to 175046)		
290	(bases 1 to 175046)		
291	(bases 1 to 175046)		
292	(bases 1 to 175046)		
293	(bases 1 to 175046)		

HOWLAND J.C., JOHNSON, R., JONES, C., KAMM, L., KARATAS, A., KLEIN, J.,  
 LANDERS, T., LEHOCZKY, J., LEVINE, R., LIU, C., LIU, G., LOCKE, K.,  
 MACDONALD, P., MARGUIS, N., MCKEAN, P., MCGURK, A., MCKERNAN, K.,  
 MCHETERS, R., MELDIRIM, J., MENDES, L., MORROW, J., NAYLOR, J.,  
 NORMAN, C.H., O'CONNOR, T., O'DONNELL, P., OLIVER, T.M., PETERSON, K.,  
 PIERRE, N., PISANI, C., POLLARA, V., RAYMOND, C., RILEY, R., ROTHMAN, D.,  
 ROY, A., SANTOS, R., SEVERY, P., SPENCER, B., STANGE-THOMANN, N.,  
 STOJANOVIC, N., SUBRAMANIAN, A., TALAMAS, J., TEELFAYE, S., THEODORE, J.,  
 TITRELL, A., VASSILIEV, H., VIEL, R., VO, A., WU, X., WYMAN, D., YE, W.J.,  
 ZIMMER, A. and ZODY, M.  
 Direct Submission  
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 175046)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Baistien, V., Beda, F.,  
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
 Campolongo, A., Castelle, A., Choepey, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, M., Gage, D.,  
 Galagan, J., Gardina, S., Glinde, S., Goylete, M., Graham, L.,  
 Girard, P., Gire, N., Grant, G., Hagos, B., Haefford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kamm, L., Karatas, A.,  
 Klein, J., Lacocque, K., Lamasz, R., Landers, T., Lehoczy, J.,  
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marguis, N.,  
 McCarthy, M., McKean, P., McGurk, A., McKernan, K., Mcheters, R.,  
 Meldirim, J., Mendes, L., Mihov, T., Miranda, C., Mleaga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Teelfaye, S., Theodore, J., Titrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-ANG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 16, 2000 this sequence version replaced gl:698033.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: U5491  
 Center clone name: 310\_J\_14  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phred, version 0.960731  
 Consensus quality: 156135 bases at least Q40  
 Consensus quality: 165373 bases at least Q30  
 Consensus quality: 168980 bases at least Q20  
 Insert size: 185000; agarose-efp  
 Insert size: 171746; sum-of-contigs  
 Quality coverage: 3.4 in Q20 bases; agarose-efp  
 Quality coverage: 3.6 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 34 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 56: contig of 56 bp in length  
 \* 57 156: gap of 100 bp  
 \* 157 1191: contig of 1035 bp in length

```
* 1192 1291: gap of 100 bp
* 1292 2432: contig of 1141 bp in length
* 2433 2532: gap of 100 bp
* 2533 3742: contig of 1210 bp in length
* 3743 3842: gap of 100 bp
* 3843 4921: contig of 1079 bp in length
* 4922 5021: gap of 100 bp
* 5022 6073: contig of 1052 bp in length
* 6074 6173: gap of 100 bp
* 6174 7552: contig of 1379 bp in length
* 7553 9001: contig of 1349 bp in length
* 9002 9101: gap of 100 bp
* 9102 10622: contig of 1521 bp in length
* 10623 10722: gap of 100 bp
* 10723 11836: contig of 1114 bp in length
* 11837 11936: gap of 100 bp
* 11937 13398: contig of 1462 bp in length
* 13399 13498: gap of 100 bp
* 13499 14949: contig of 1451 bp in length
* 14950 15049: gap of 100 bp
* 15050 16260: contig of 1211 bp in length
* 16261 16360: gap of 100 bp
* 16361 18265: contig of 1905 bp in length
* 18266 18365: gap of 100 bp
* 18366 19916: contig of 1551 bp in length
* 19917 20016: gap of 100 bp
* 20017 22939: contig of 2923 bp in length
* 22940 23039: gap of 100 bp
* 23040 26445: contig of 3406 bp in length
* 26446 26545: gap of 100 bp
* 26546 29327: contig of 2782 bp in length
* 29328 29427: gap of 100 bp
* 29428 32200: contig of 2773 bp in length
* 32201 32300: gap of 100 bp
* 32301 36738: contig of 4438 bp in length
* 36739 36838: gap of 100 bp
* 36839 43447: contig of 6609 bp in length
* 43448 43547: gap of 100 bp
* 43548 48525: contig of 4978 bp in length
* 48526 48625: gap of 100 bp
* 48626 52791: contig of 4166 bp in length
* 52792 52891: gap of 100 bp
* 52892 58391: contig of 5500 bp in length
* 58392 58491: gap of 100 bp
* 58492 64869: contig of 6378 bp in length
* 64870 64969: gap of 100 bp
* 64970 73369: contig of 8400 bp in length
* 73370 73469: gap of 100 bp
* 73470 81101: contig of 7632 bp in length
* 81102 81201: gap of 100 bp
* 81202 87754: contig of 6553 bp in length
* 87755 87854: gap of 100 bp
* 87855 95534: contig of 7680 bp in length
* 95535 95634: gap of 100 bp
* 95635 102797: contig of 7153 bp in length
* 102798 102897: gap of 100 bp
* 102898 111817: contig of 8920 bp in length
* 111818 111917: gap of 100 bp
* 111918 121160: contig of 9243 bp in length
* 121161 121260: gap of 100 bp
* 121261 140537: contig of 19277 bp in length
* 140538 140637: gap of 100 bp
* 140638 175046: contig of 34409 bp in length.
Location/Qualifiers
1. 175046
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="1"
/clone="RP11-310J14"
/clone_lib="RP11 Human Male BAC"
```

```
misc_feature 1..56
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 157..1191
/note="assembly_fragment"
1292..2432
/note="assembly_fragment"
2533..3742
/note="assembly_fragment"
misc_feature 2533..3742
/note="assembly_fragment"
3843..4921
/note="assembly_fragment"
5022..6073
/note="assembly_fragment"
misc_feature 5022..6073
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature 6174..7552
/note="assembly_fragment"
7553..9001
/note="assembly_fragment"
9102..10622
/note="assembly_fragment"
10723..11836
/note="assembly_fragment"
11937..13398
/note="assembly_fragment"
13499..14949
/note="assembly_fragment"
15050..16260
/note="assembly_fragment"
16361..18265
/note="assembly_fragment"
misc_feature 59.9%; Score 505; DB 2; Length 175046;
Best Local Similarity 71.8%; Pred. No. 5e-94;
Matches 610; Conservative 0; Mismatches 233; Indels 6; Gaps 3;
Query 1 GGGCAAAANNTAATTTCTTANCTCCTGCAAAANCTATTATTTACCC 60
Db 43358 GGGCAAAAGTTATTTTATTTATTAACCTTCTGCAAGCATTATTTATCC 43299
61 TATNNCNCNGANTTANAAATACCTTNTTTAAAAAACCCTNGAAAAATTA 120
Db 43298 TATATCTACGAAATTTTAAGAAATACATTAGATTGAAAAAAGATA 43239
121 TNGCAATPANTTAACTTCTTGAANAGAAATTTTACCAAGACNGAANCTTNT 180
Db 43238 ATGCAGATPAAATTAACCTTACGAAAAAGAAATTTATACAAAGACTGAGACGTAT 43179
181 AATNGAANTAAATTAATANTNGAANCGGCGNNGAACAACCTTATGTCATT 240
Db 43178 AAATGAAATAGATTAATTAATTTGAAACCTGATCGAAGCAAACTTTATGTCATT 43119
241 ATCTTAANGAGGNNTTTANNACTAATNCCGATTTTCCATANGAANCCNNNT 300
Db 43118 ATCTTAATGATGGGTTTATGACTATATACGATTTTTCANTAAGAAACCATGT 43059
301 AAAANTTTTNTTAAATTAATAAACCNGTCCACCCCGATCANATTCCTTNA 360
Db 43058 AAAATATTTTATTTTAAATAAAGCTGTTCAAGCTGATCATATTTCTTTAT 42999
361 TGGATGGGAAAAAANCCNGTCCNNATACCNNGAANNGAANCTTTAAATTT 420
Db 42998 TTGATTTGGGAAAGAAATACGTTTCTGATGCAATGAATGCAAAATTTTGA 42939
421 ACCCCCTANTTTTAAANCTATNGAANAANTNGATTGGAATTCG-CAACCTTA 479
Db 42938 ATCTCAGTAATTTTAAAGAACTATTTGAGAAATTTGATTAATGACATGACAACTTA 42819
480 NTTCNGGCAACNGTGGCCTGCTTCTTACTTANTTCCCCCAAGAAANCTTAA 539
Db 42818 ATTACTGCGCAGCTGTTGGCATGTTGTTCTTACTTAATTTCTCCCAAGAAACCTTAA 42819
```

Qy 540 NCNGAANTCCNCCAAATACCTTANTATCTTGTAACCAAAACCTTTTN 599  
 Db 42818 ACTGAATCTTCAGCAGAAATACCTTAATATCTTGTAACCAAAACCTTTTN 42759  
 Qy 600 GTTACNTANTCTTGGGATTTAACGGGTCCCAATTNATCCGAAACCCATTTTCCC 659  
 Db 42758 GTTACATAGTCTTGGATTTTACGTTCTAATTTATCTGAAACCAATTTTACC 42699  
 Qy 660 CNAACCATANTTACATTTTACCTTGTAAGCNCAGTGTGCAANTCCGCAANCG 719  
 Db 42698 CAGACCATATATACATATTAATCTTGTAATGACAGTGTGCAATTCGCAAGCAG 42639  
 Qy 720 TANTTTTCC--CNGGNCCTTCCCGGNCCTTGGAAGAAACGGGATNGTCCCGCT 777  
 Db 42638 TAGTATACCATAGGCTCTATTCACCCAGCATAGGAAACAGATTGGTACACCA 42579  
 Qy 778 T---AAAAACAACCTTCCCGCCCTTGCCAGGNTTNTTCCGTTAAATCCGAAC 834  
 Db 42578 TTAAGAAACAATCTATCCCAAGCTATTTGCCAATGTGTAGTTTCACTAATCTGAC 42519  
 Qy 835 AATAAAG 843  
 Db 42518 AATAAAG 42510

RESULT 6  
 LOCUS AL445183/c 193774 bp DNA linear PRI 05-APR-2002  
 DEFINITION Human DNA sequence from clone RPI1-334A14 on chromosome 1, complete  
 ACCESSION AL445183  
 VERSION AL445183  
 KEYWORDS HTG. GI:20068427  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 AUTHORS Wallis, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humgeny@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Apr 7, 2002 this sequence version replaced gi:17939714.  
 COMMENT  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr1  
 RPI1-334A14 is from the library RPI1-11.2 constructed by the group  
 of Pleyer de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6.

FEATURES  
 Source  
 1..193774  
 Location/Qualifiers  
 /organism="Homo sapiens"

/mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RPI1-334A14"  
 /clone\_1ib="RPI1-11.2"  
 BASE COUNT 50815 a 45351 c 44031 g 53577 t  
 ORIGIN

Query Match 59.9%; Score 505; DB 9; Length 193774;  
 Best Local Similarity 71.8%; Pred. No. 5e-94;  
 Matches 610; Conservative 0; Mismatches 233; Indels 6; Gaps 3;

Qy 1 GGGCAAAAAAATTTTAAATTTCTTAAATCTTCCGCAAAACCTTTTACC 60  
 Db 98225 GGGCAAAAAAATTTTAAATTTCTTAAATCTTCCGCAAAACCTTTTACC 98166  
 Qy 61 TATNNCNCNGANTTTNANAANTACCTTNTNTTAAAAAACCCTGAAAAAATTA 120  
 Db 98165 TATATCTCAGTAATTTTAAAGAAATACATTAGTATTGAAAAAATAGATA 98106  
 Qy 121 TNGCAATANTTACCTTCTTGAAGAAATTTTACCAAGACGAAACCTTNT 180  
 Db 98105 ATGCAGATTAATTAACCTTACGAAAAAAGAAATTAACAAAGACTGAGACGTTAT 98046  
 Qy 181 AATTNGAANTAAATTAATTTNGAAGGCGGNCNGAAGCAACCACTTATGTCAT 240  
 Db 98045 AAATGAAATGAGATTAATTAATTTGAAACCTGATCTGAAGCAACCTTATGTCAT 97986  
 Qy 241 ATCCCTNANGAGGNNTTNANNACTAATNCCGATTTTCCATANGAANCCNNNT 300  
 Db 97985 ATTCTTAATGATGGTGTTTATGACTAATACGATTTTCAATAGGAAACCATGT 97926  
 Qy 301 AAAANTTTTNTAATTTAAAAATTAACCCGNTTCAACCCCGATCANATCTTTTAT 360  
 Db 97925 AAAATATTTTATTTAAAAATTAAGCCTGTGTTCAACCTGATCATATTTCTTTAT 97866  
 Qy 361 TGGATTTGGGAAAAAATNCCGTTCCNNAATCCNNGAANNGCAANTTTTAAATTTTA 420  
 Db 97865 TTGATTTGGGAAAAAATACGTTTCTGATGACATGAATGCAAAATTTTGAATTTTA 97806  
 Qy 421 ACCCCCTAATTTTAAAAAANTATNGAANAANTNGATTANNGACTGAATTCG-CAACCTTA 479  
 Db 97805 ATCTCAGTAATTTTAAAGAACTATTAAGAAATTTGATTAATGACATGAAGCAACACTTA 97746  
 Qy 480 NTTCNGGCCACGNGTGGCCTNTGTTCTTACTTANTTCCCGCAAGAAANCCCTTA 539  
 Db 97745 ATTACTGGCCAGCTGTGTCATTTGTTCTTACTTACTTCCCAAGAAAACTCTTA 97686  
 Qy 540 NCNGAANTCCNCCAAATTAACCTTANTATCTTGTAACCAAAACCTTTTN 599  
 Db 97685 ACTGAATCTTCAGCAGAAATACCTTAATATCTTGTAACCAAAACCTTTTN 97626  
 Qy 600 GTTACNTANTCTTGGGATTTAACGGGTCCCAATTNATCCGAAACCCATTTTCCC 659  
 Db 97625 GTTACATAGTCTTGGATTTTACGTTCTAATTTATCTGAAACCAATTTTACC 97566  
 Qy 660 CNAACCATANTTACATTTTACCTTGTAAGCNCAGTGTGCAANTCCGCAANCG 719  
 Db 97565 CAGACCATATATACATATTAATCTTGTAATGACAGTGTGCAATTCGCAAGCAG 97506  
 Qy 720 TANTTTTCC--CNGGNCCTTCCCGGNCCTTGGAAGAAACGGGATNGTCCCGCT 777  
 Db 97505 TAGTATACCATAGGCTCTATTCACCCAGCATAGGAAACAGATTGGTACACCA 97446  
 Qy 778 T---AAAAACAACCTTCCCGCCCTTGCCAGGNTTNTTCCGTTAAATCCGAAC 834  
 Db 97445 TTAAGAAACAATCTATCCCAAGCTATTTGCCAATGTGTAGTTTCACTAATCTGAC 97386  
 Qy 835 AATAAAG 843  
 Db 97385 AATAAAG 97377

RESULT 7  
HUMSCP2A/c 2572 bp mRNA linear PRI 06-DEC-1993  
LOCUS Human sterol carrier protein X/sterol carrier protein 2 mRNA,  
DEFINITION complete cds.  
ACCESSION M75883  
VERSION M75883.1 GI:432974  
KEYWORDS sterol carrier protein-2, sterol carrier protein X.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 2572)  
AUTHORS He,Z., Yamamoto,R., Furch,E.E., Schantz,L.J., Naylor,S.L.,  
George,H., Billheimer,J.T., and Straus,J.F. III.  
CDNA encoding members of a family of proteins related to human  
sterol carrier protein 2 and assignment of the gene to human  
chromosome 1 p21---pter  
JOURNAL DNA Cell Biol. 10 (8), 559-569 (1991)  
MEDLINE 92029618  
PUBMED 1718316  
REFERENCE 2 (bases 1 to 2572)  
AUTHORS Vesa,J., Hellsten,E., Emanuel,B.S., Billheimer,J.T.,  
Mead,S., Cowell,J.K., Straus,J.F. III., and Peltonen,L.  
Assignment of sterol carrier protein X/sterol carrier protein 2 to  
1p33 and exclusion as the causative gene for infantile neuronal  
ceroid lipofuscinosis  
JOURNAL Unpublished  
COMMENT On Dec 6, 1993 this sequence version replaced gi:410029.  
ORIGINAL SOURCE text: Human liver cDNA to mRNA.  
FEATURES  
source 1..2572  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="1p32"  
/tissue\_type="liver"  
/clone\_id="Clontech HL-1001b;HL-1115b"  
1..2572  
/gene="SCP-X/SCP-2"  
22..1665  
/gene="SCP-X/SCP-2"  
/codon\_start=1  
/product="sterol carrier protein X/sterol carrier  
protein-2"  
/protein\_id="AAA03557.1"  
/db\_xref="GI:432975"  
/translation="MSSSPWEPATLRVFEVGVGKTPVKGPAENSRDYPDLAEBAK  
KLAADAIPIYSAVDACVGVFGDSTCGORAIYHSLGTMGDIPIVNNNCATGSTALF  
MAROLIQGVAVACVLAIEPKMSKSLGKPSDRITPDKHDVLLINXGSAHVAP  
OMFYAGKHEKRYGTIKIEHFAKIKGKHKHSHVNNPISPOFDEYSLDEVMASKEVDF  
LTLLOCCTSDGAAAILASSEAFOVKYGIQSAVEILAQEMMTDLPSSFEKSLIKMV  
GPDMSKEARKCYEKSGLPNDIDVIEHDCFTNELLYEALGLCPDEOGATLVDRG  
DNTYGGKWIINSGGLISKHGLATGALCAELCMQLRGEAGKQVPAKVALQHL  
GIGAVVVTLLYKMGPEPAASSFRTHQIEVNGSSASPSDKALVKEIEKKEEGEO  
PVKIKGIFAEKVDGPGKKEATWVVDVNGKSVLPNSDKRADCTITMADSDPLALM  
TGRKMPQSAFFQGGKLIKTNMGLAMKQNLQLOPNAKU"  
polyA\_signal 1932..1937  
/gene="SCP-X/SCP-2"  
/evidence=not\_experimental  
polyA\_signal 2213..2218  
/gene="SCP-X/SCP-2"  
/evidence=not\_experimental  
polyA\_signal 2258..2263  
/gene="SCP-X/SCP-2"  
/evidence=not\_experimental  
polyA\_signal 2341..2346  
/gene="SCP-X/SCP-2"  
/evidence=not\_experimental  
polyA\_signal 2550..2555  
/gene="SCP-X/SCP-2"  
/evidence=experimental  
polyA\_site 2572  
/evidence=experimental

BASE COUNT 794 a 439 c 570 g 769 t  
ORIGIN  
Query Match 59.8%; Score 504; DB 9; Length 2572;  
Best Local Similarity 71.8%; Pred. No. 1.5e-93;  
Matches 609; Conservative 0; Mismatches 233; Indels 6; Gaps 3;  
/gene="SCP-X/SCP-2"  
/evidence=experimental  
2 GCCAAAAAANTATTNAAATTTCTATTANCTCTCCNCAANCAATTNACCT 61  
2571 GCCAGAAAAAGTATTATTTTCAATTTCAATTTCAATTTCAATTTATCT 2512  
QY 62 ANNMCNCGNANTTTNAAANTACCTTNNNTTAAAAACNNGAAAAATAAT 121  
DB 2511 AATCTCACTGATTTTAAAGAAATACATGTGATTAAGAAAACTAAGAAAAAGATAA 2452  
QY 122 NGCAATANTTAACTTCTTGAAGAAATTTTACCAAGACGCAANCAANTTTA 181  
DB 2451 TGCAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2392  
QY 182 ATTNGAANTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 241  
DB 2391 AATTGAAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2332  
QY 242 TCCCTAAGAGGNNNTTANNAACATNCCNGATTTTCAATNAGGAAACCATTTA 301  
DB 2331 TTCTTAAGATGATGTTTATTAAGATTAATTAATTAATTAATTAATTAATTAATTA 2272  
QY 302 AAANTNTTTTAAATTTAAATTAATTAATTAATTAATTAATTAATTAATTT 361  
DB 2271 AAAATATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTTCTTTATTT 2212  
QY 362 GGATTGGGAAAAATNCGTTCCNATPACCNNAANNGCAANTTTTAAATTTTAA 421  
DB 2211 TGATTTGGGAAAGAAATCTGTTCTGATAGATGAATGAATTTTATGATTTTAA 2152  
QY 422 CCCCCCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 480  
DB 2151 TCTCACTAATTTTAAAGAACTATTAAGAAATTAATTAATTAATTAATTAATTA 2092  
QY 481 TTNCNGCCACCGNGTGGGCTNNTTCTTACTTAATTCCTCCCAAGGAAANCTTAN 540  
DB 2091 TTACTGGCAGCTGTGGCATTTGTCTTTACTTACTTCTCCCAAGGAAANCTTAA 2032  
QY 541 CNGAATCTCCNCAAAATPAACCTTAAATTAATTAATTAATTAATTAATTTTNG 600  
DB 2031 CTGAATCTTCAGCAAAATPAATCTTAAATTAATTAATTAATTAATTAATTTTG 1972  
QY 601 TTTAATNTAATCTTGGGATTTTAAAGGCTCCCAATTAATTAATTAATTAATTTCC 660  
DB 1971 TTTAATATGTTCTTTGGATTTTAAAGGCTCTTAATTTTCTGAAATCTCAATTTA 1912  
QY 661 NAACATANTTAACATTTTAACTTGGTGAAGCCNCAAGTGTTCANTNCCGAANCACT 720  
DB 1911 AGACCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1852  
QY 721 AATNTTCC--CCNGCCTTTCCCGGACNCTTGGGAAAAACGGATNGTCCCTT 778  
DB 1851 AGTATACATCAGGCTCTATTCACCCAGAGATAGGAAAAACAGATGTGTACACCAT 1792  
QY 779 ---AAAAACAACCTTCCCGCCTTTGGCCAGGNNNTTTCCTTAATTCGAACA 835  
DB 1791 TAAAGAAACAAATCTTATCAGCTATTGTCATGTTGAGTTTCAATTAATTTGACA 1732  
QY 836 ATAAAAAG 843  
DB 1731 ATAAAAAG 1724  
RESULT 8  
BC005911/c  
LOCUS BC005911 1439 bp mRNA linear PRI 12-JUL-2001







```

RESULT 11
ARI46580 LOCUS ARI46580 687 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 20 from patent US 6218521.
ACCESSION ARI46580
VERSION ARI46580.1 GI:15109769
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 687)
TITLE Obata, Y.
JOURNAL Isolated nucleic acid molecules associated with gastric cancer and
FEATURES method for diagnosing and treating gastric cancer
PATENT: US 6218521-A 20 17-APR-2001;
LOCATION/Qualifiers
source 1..687
BASE COUNT 242 a 98 c 85 g 258 t 4 others
ORIGIN
Query Match 48.8%; Score 411.8; DB 6; Length 687;
Best Local Similarity 71.8%; Pred. No. 1.5e-74;
Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;

OY 15 ATTNAATTTCTTATTANCCCTTCGCAACCATTAATTATTCCTATNNCNCNGAN 74
DB 1 ATTNAATTTCTTATTAAACATTTCTCTCAAGCATTAATTATTCCTATATCTCACTGA 60
OY 75 TTTNNAANAATACCTTNTNTTAAACCTNCGAATAAATAATNCAATNTTAA 134
DB 61 TTTTANNAATAATACATTTAGTATTAGAAAACATGAGAAAAGATNAATGACATATTTAA 120
OY 135 CCTTCTGAAAANGAATAATTNTTACCAANGAGCAAGAAACNTTAAATTNGAATNAAA 194
DB 121 ACTTACATGAAAAGAAAATTTATTAACAAGAGACTGAGAACCTTATTAATTGAAATGACA 180
OY 195 TTATANTNGAANGCGGCGNCGAAGAACCAANTTNATGGTCCAAATTAATCTTANAGAGG 254
DB 181 TTATATATTGAAAACATGCAATCTGAAAGCAAACTTATATGTTCAATATCTTAATGATGG 240
OY 255 NNTTNNANNACTTAATNCCNGATTTTCCAAATANGAAGAACCCNNNTTAAANNTTTNAT 314
DB 241 TGTTTATGACTAAATACACTGATTTTTCATTAAGGAAACCATGTTAATAATATTTTAT 300
OY 315 TTTAANAATTAACCCGNTTCCAAACCCGATCANATTCCTTNNATTTGATGGGAAAA 374
DB 301 TTTTAAAATAATAGCCTGTGTGTTCAAGCTCTGATCATATTTTATTTATTTGATGGAGAA 360
OY 375 AAATNCGTTCNNATACCCNNGAANNCAAAATTTTAAATTTTAAACCCCTATNTTT 434
DB 361 AAATACGTGTTCTGATAGATGAAATGCAAAATTTTATGATTTTAAATCTCCTATTTT 420
OY 435 AAAAATCTAATGAAANTNGATTAATNGACTTGAATGCG-CAACCTTANTTNCNGGCCACN 493
DB 421 AAGAACTATGGAAGAAATGATTAATGACATGAAGTGCACAAACATTAATTAATGCGCAGT 480
OY 494 GTGGCGNTNGTTCCTTACTTANTCCGCCAAGGAANNCTTANNGAANGACTCCNCC 553
DB 461 GTTGGCATTTGTCTTTCTTACTTATGTTCTCCCAAGGAAAACCTTTAACTGAATCTTCACG 540
OY 554 AAAATACCTTAATATCTTGTGTAACCAANCAAAACCTTTTNGTTTACNTATNTCT 613
DB 541 NGAATTAACCTTAATATATCTTGTAGCCAAAC--AAAACCTTTTGTGTTACATAGTCT 598
OY 614 TGGGATTTAAAGGGTCCCAATTTTATCGGAAGCAACCATTTTCCCGCAACCATANTTAC 673
DB 599 TTGGATTTTACATGTTCTTAATTTTATTCGAAAGCTCAATTTTCCCGACACATATATAC 658
OY 674 CAT-TTAACTTGTAAAGGCGAGTNGTT 701

```

```

DB 659 CCTATTAACTTTGTTATGACAGTGTGT 687

RESULT 12
BD079829 LOCUS BD079829 687 bp DNA linear PAT 27-AUG-2002
DEFINITION Cancer-associated nucleic acids and polypeptides.
ACCESSION BD079829
VERSION BD079829.1 GI:22625432
KEYWORDS JP 2001516009-A/495.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 687)
JOURNAL Old, L. J., Scanlan, M. J., Stockert, E., Gure, A., Chen, Y. T., Gout, I.,
Ognare, M., Obata, Y., Pfeundscher, M., Tureci, O., and Sahin, U.
Cancer-associated nucleic acids and polypeptides
PATENT: JP 2001516009-A 495 25-SEP-2001;
LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT OS Homo sapiens (human)
PN JP 2001516009-A/495
PD 25-SEP-2001
PF 15-JUL-1998 JP 2000503425
PR 17-JUL-1997 US 08/896164, 10-OCT-1997 US 60/061599 PR
10-OCT-1997 US 60/061765, 10-OCT-1997 US 08/948705 PR
11-OCT-1997 GB 9721697.2, 22-JUN-1998 US 09/102322 PI
J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI
CHEN,
PI IVAN GOUT, MICHAEL O'HARE, YUICHI OBATA, MICHAEL PFEUNDSCHUH, PI
OZLEM TURECI,
PI UGUR SAHIN
PC G01N33/574, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61K48/00, PC
A6P35/00,
PC C07K14/42, C07K16/32, C12N15/09//C07K16/46, C12P21/08, A61K37/02,
PC C12N15/00
CC Cancer-associated nucleic acids and polypeptides. FH Key
LOCATION/Qualifiers
FT source 1..687
FT Location/Qualifiers
FEATURES 1..687
source /organism='Homo sapiens (human)'.
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 242 a 98 c 85 g 258 t 4 others
ORIGIN
Query Match 48.8%; Score 411.8; DB 6; Length 687;
Best Local Similarity 71.8%; Pred. No. 1.5e-74;
Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;

OY 15 ATTNAATTTCTTATTANCCCTTCGCAACCATTAATTATTCCTATNNCNCNGAN 74
DB 1 ATTNAATTTCTTATTAAACATTTCTCTCAAGCATTAATTATTCCTATATCTCACTGA 60
OY 75 TTTNNAANAATACCTTNTNTTAAACCTNCGAATAAATAATNCAATNTTAA 134
DB 61 TTTTANNAATAATACATTTAGTATTAGAAAACATGAGAAAAGATNAATGACATATTTAA 120
OY 135 CCTTCTGAAAANGAATAATTNTTACCAANGAGCAAGAAACNTTAAATTNGAATNAAA 194
DB 121 ACTTACATGAAAAGAAAATTTATTAACAAGAGACTGAGAACCTTATTAATGGAATGAGA 180
OY 195 TTATANTNGAANGCGGCGNCGAAGAACCAANTTNATGGTCCAAATTAATCTTANAGAGG 254
DB 181 TTATATATTGAAAACATGCAATCTGAAAGCAAACTTATATGTTCAATATTTCTTAATGATGG 240
OY 255 NNTTNNANNACTTAATNCCNGATTTTCCAAATANGAAGAACCCNNNTTAAANNTTTNAT 314
DB 241 TGTTTATGACTAAATACACTGATTTTTCATTAAGGAAACCATGTTAATAATATTTTAT 300

```



COMMENT On Dec 6, 1993 this sequence version replaced gi:432972.  
Original source text: Human female liver.

## FEATURES

source  
1..1219  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="1p32"  
/tissue\_type="liver"  
1..1219  
/gene="SCP-2"  
84..515  
/gene="SCP-2"  
/codon\_start=1  
/product="sterol carrier protein-2"  
/protein\_id="AA03559.1"  
/db\_xref="GI:432973"  
/translation="MGPEAASSFRTHOIEAVPTSSASDPKANIIVFEIEKLEEG  
EOPVKIGIGIFAFKVDGKGEATWVDVKKGKSLVPSNDKKADCTTMADSDFLA  
LMTGKNPQSAFQGLKITGTMGLAMKLQNLQDPNAKL"  
84..143  
/gene="SCP-2"  
144..512  
/gene="SCP-2"  
/product="unnamed"  
782..787  
/gene="SCP-2"  
/evidence="not\_experimental"  
1062..1067  
/gene="SCP-2"  
/evidence="not\_experimental"  
1107..1112  
/gene="SCP-2"  
/evidence="not\_experimental"  
1190..1195  
/gene="SCP-2"  
/evidence="not\_experimental"  
1219  
/evidence="experimental"  
/gene="SCP-2"  
/evidence="experimental"  
BASE COUNT 402 a 195 c 255 g 367 t  
ORIGIN

Query Match 42.3%; Score 357; DB 9; Length 1219;  
Best Local Similarity 71.4%; Pred. No. 2.5e-63;  
Matches 461; Conservative 0; Mismatches 178; Indels 7; Gaps 4;

204 GAANAGCGGNNNGAAGAACCAACCACTTATGTCCTCAATTAATCTTAANAGAGGNNTTTANN 263  
1218 GAAAGCTGCATCTGAAAGCAACCTTATTGTTCAATTTCTTAATGATGCTTTATG 1159  
264 ACTAATNCCGAGATTTTCCATATANGAANCCNNNTTAAANTTTTATTTTAAAT 323  
1158 ACTAATACCTGATTTTCAATTAAGAAACCATGTTAAATAATTTTATTTTAAAT 1099  
324 AACCGNRTTCAACCCCGATCAAAATTCCTTMAATTTGGATTTGGGAAAAAATNCGT 383  
1098 AAGCTGTGTTCACTGATCATATTTCTTTATTTTGAATTTGGGAAAAAATACGT 1039  
384 TCCNNAATACNNNGAANNNGCAANTTTTAAATTTTAAACCCCTANTTTTAAANCT 443  
1038 TTCTATATGACATGAATGCAAAATTTTATGATTTT-ATCTACATTAATTTTAAAGACT 980  
444 NGAAANANTNGATTANNGACTGAATTGC-CAACCTANTTNGCGCAACNGTGGCUNT 502  
979 TGAGAAATATGATTAATGACATGAGTGCACACCTAATTAATCTGCGCAGCTGTTGGCATT 920  
503 GTTTCCTTACTTANTCCCCCAAGAAANNCCTTAANCGAANTCTCCNCCAAATTAAC 562  
919 GTGTTTCTTACTTACTTCTCCCAAGAAACCTTAACCTGAATCTTCGACGAATTAAC 860  
563 CTTAANTATCTTGGTAACCAANCAAACTTTTNGTTTGCNTAANTCTTGGGATTTA 622

Db 859 CTTAATAATATCTTGTGAAGCAAAACAAAGCTTTTGTGTACATAGTCTTTGGATTTT 800  
Qy 623 ACGGTCCCCCAATTTATCCNAGAACCCANTTTTCCCCCAACCATANTTACATTTTACC 682  
Db 799 ACTGTTCTTAATTTTATTTCTGAACTCAATTTTACCAGACCATTAATTTACATATTAAC 740  
Qy 683 TTGTGAGGNCAGTNGTTTGCANTNCGCAANAGTANTTTTCC--CCNGCNCCTTTC 740  
Db 739 TTTGTAATGACAGTGTATGCAATTTCCGCAAGAGTAGTATACATCAGGCTCTATTTC 680  
Qy 741 CCCCAGNCTTGGGAAAAACGGATNGTCCCTCCCTT---AAAAACAACCTTCCCN 797  
Db 679 ACCCAGACATATGAAAAACAGGATTTGTCACACCATTAAGAAACAAATCTATCCAG 620  
Qy 798 CTTTGGCCGACAGNTTNTTCCCTTAATCCGAACATATAAAG 843  
Db 619 CTATTTGCCAATGTGTAGTTTCAGTCTAAATTTGACATATAAAG 574

RESULT 15  
HUMSCP2B/c 1500 bp mRNA linear PRI 06-DEC-1993  
LOCUS Human sterol carrier protein 2 mRNA, complete cds.  
DEFINITION M75884  
ACCESSION M75884.1 GI:432976  
VERSION  
KEYWORDS sterol carrier protein-2.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 1500)  
He,Z., Yamamoto,R., Furch,E.E., Schartz,L.J., Naylor,S.L.,  
George,H., Billheimer,J.T., and Straus,J.F., III.  
cDNAs encoding members of a family of proteins related to human  
sterol carrier protein 2 and assignment of the gene to human  
chromosome 1 p21---pter  
DNA Cell Biol. 10 (8), 559-569 (1991)  
92029618  
MEDLINE 1718316  
PUBMED 2 (bases 1 to 1500)  
Vees,J., Hellsten,E., Branoski,B.L., Emanuel,B.S., Billheimer,J.T.,  
Mead,S., Cowell,J.K., Straus,J.F., III.  
Assignment of sterol carrier protein X/sterol carrier protein 2 to  
1p32 and exclusion as the causative gene for infantile neuronal  
ceroid lipofusidosis  
Unpublished  
JOURNAL On Dec 6, 1993 this sequence version replaced gi:337996.  
COMMENT Original source text: Human liver cDNA to mRNA.  
FEATURES  
source  
1..1500  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="1p32"  
/tissue\_type="liver"  
/clone\_lib="Clontech HL-1001b;HL-1115b"  
1..1500  
/gene="SCP-2"  
82..951  
/gene="SCP-2"  
/codon\_start=1  
/product="sterol carrier protein-2"  
/protein\_id="AA03558.1"  
/db\_xref="GI:432977"  
/translation="MTDLPSSFEKSIIRKVPFMSKRAKCYEKSGLTENDIDVI  
ELHOCFSNELITVEALGLCPGGGATLVDGDNTRYGKWVNPSSGLISKGPFGAT  
GLAACRLCMOLRGEAGKRVPGAAVLAQNLGIGATVYVTLTKRFPBAASSFRTHO  
IEAVPTSSASDPKANIIVFEIEKLEEGEOPVKIGIGIFAFKVDGKGEATWVDVKKGKSLVPSNDKKADCTTMADSDFLA  
LMTGKNPQSAFQGLKITGTMGLAMKLQNLQDPNAKL"  
1218..1223  
/gene="SCP-2"  
/note="no polyA or polyA signal was found at the end of

this cDNA clone, indicating that the mRNA sequence may be  
partial at the 3' end  
/evidence=not\_experimental

BASE COUNT 482 a 254 c 342 g 422 t  
ORIGIN

Query Match 31.2%; Score 262.8; DB 9; Length 1500;

Best Local Similarity 70.9%; Pred. No. 5.2e-44;  
Matches 348; Conservative 0; Mismatches 137; Indels 6; Gaps 3;

```
QY 359 TTGGATTGGGGAAAAAATNCNGTCCNNATACNNNGAANNCAANTTTTAAATTT 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1500 TTTGATTGGGAGAGAAATACGTTCTGATAGCATGAATGCAAAATTTTAGATTT 1441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 419 TAAACCCCTTATTTTAAANCTAINGAAAANTNGATTANNGACTTGAAATGC-CAACCC 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1440 TAATCTCACTAATTTTAAAGAACTATGAGAAATGTATATGACATGAAGTGACAAAC 1381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 478 TANTTNCNGGCCACCGTGGGCTNTGTTCTTACTTANTCCCCCAAGAAANNCTT 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1380 TAATTAAGTGGCCAGCTGTGGCATGTGTTCTTACTTACTGTTCTCCCAAGAAACTCTT 1321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 538 AANCGAANCCTCCNCCAAATTAACCTTANTATCCTGTGTAAACCAANCAAAACCTTT 597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1320 AAACGAATCTTCAGCAGAAATATCTTAAATATCTTGTAAAGCAAAACAAAGCTTT 1261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 598 TNGTTTACNTANTCCTTGGGATTTAACGGGTCCCAATTNATCCNGAACCCANTTTCC 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1260 TTGTTTACATAGTCTTGTGATTTTACTGTCTCTAATTTTATCTGAACCTCAATTTAC 1201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 658 CCCNAACCATANTTACATTTTACCTTGTGAAGCNCAGTNGTTCANTNCCGAAANC 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1200 CCCAGCCCAATTAATACATTTAATCTTGTAAATGACAGTGTATGCAATTCGCAAAAGC 1141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 718 AGTANTNTTCC--CCNGGNCCTTCCCCCGANCTTGGGAAAAAACGGATNGGTCCCCC 775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1140 AGTAGATATCCATCAGGCTCTATTACCCAGAGCATAGAAAAACAGATTGTCACACC 1081
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 776 CTT--AAAAACAACCTTCCCCCNCCTTGGCCCGAGNNNTTNTCCGCTTAATCGA 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1080 CATTAAGAAACAATCTTAATCCAGCTATTGCGCAATGTAGATTCAAGCTAAATTTCTG 1021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 833 ACAATTAAG 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1020 ACAATTAAG 1010
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: November 27, 2003, 10:02:41  
Job time : 3416.44 secs







Oy		201	TTCGAAACGGCCNNNGAAGCAACTNTNATGTCTCATTTACCNTANANGAGGNATTN	260
Db		183	TTTGAATACTGCATCTGAAAAGAACAATTATATGTTCAAATATCTTAATGAAGTGTTTT	242
Oy		261	AANNACTTAATNCSCNGATTTTCCAATFANGSAAANCSCNNNTTAAANTNTTTAATTTTAA	320
Db		243	ATGACTATATACACTGATTTTTCATATAGSAAACCATGTATAAAAATATTTTATTTTAA	302
Oy		321	AATAAACCGNTNTCCAACCCCNGATCANATTCCTTTNAITTTGGATTGGGAAAAAAAAATNC	380
Db		303	AATAAGCCTGTTTTCAAGCTGCATCATATTTCTTTTATTTTGGATTTGGSAAGAAAAATAC	362
Oy		381	NGTTCCCNNAATACCCNNAANGCAANTTTTAAATTTTAAACCCCCCTANTTTTAAANC	440
Db		363	TGTTTTCTGATAGCAGTAATGCMAAATTTTATGATTTTAACTCATATATTTTAAAGAAC	422
Oy		441	TATNGAAANAATNGATTAAINGACTGAAATTC-CAACCCCTANTTNNGGCCACNGTGGGC	499
Db		423	TATTAGAAATTTGAATTAATGACATGACATGACAACTAATTAATCTGGCCAGCTTTGGC	482
Oy		500	NTEGNTTTCCTTACTTANTTCCCCCCAAGSAAANNCCTTAAMCSGANCTCCNCAAAAATA	559
Db		483	ATGTHTTTTCTTACTAGTGTCTCCAAAGSAAACCTTAATACTGATCTTCAGCAGAATA	542
Oy		560	ACCCTTAANTATCTTGGTGAACCAAAACAAACCTTTTNGTTTACNTANTCTTGGAT	619
Db		543	ATCCTTAATAATATCTTTGTGACCMMAACMAAGCTTTTGTTCATATAGTCTTTGGAT	602
Oy		620	TTAAGSGGGTCCCAATTTNATCCNGMACCCANTTTTCCCCCAACCATANTTACCATTTT	679
Db		603	TTTACTGTTCCTTAATTTTATTTCTGAATCTCAATTTTACC CGANCATATATTAACATAT	662
Oy		680	ACCTTGTGAAGCNCAGNTNGTTGCANTNCCGCAANCAGTANTTTC--CCNGSCNCT	737
Db		663	AAC TTGTATATGCACAGTTGTATGCAATTYGCAAGAGCATGTATACATCAGGCTGTA	722
Oy		738	TTCCCCCGCAGCTTGGGAAAAACGGATNGTGTCCCCCTT---AAAAACAACCTTCCC	794
Db		723	TTCACCCAGAGCATATGAAAAACAGATTTGTCCACACCCATTATAGAAACMAATCTATCCC	782
Oy		795	CNCCTTTGGCCACAGNNTNTTCCCGCTAATCGAACATATAAAG	843
Db		783	ACGCTATTTGCCAATGTGTAGTTTGATCTAATTTGTGCAATPAAAAATg	831
RESULT_3 BX392565				
LOCUS DEFINITION	BX392565 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS01022YFOI 3-PRIME, mRNA sequence.	901 bp mRNA linear EST 13-MAY-2001		
ACCESSION VERSION KEYWORDS	BX392565 BX392565.1 GI:30607982 EST.			
SOURCE ORGANISM	Homo sapiens (human)			
REFERENCE AUTHORS TITLE JOURNAL COMMENT	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo. 1 (bases 1 to 901) L.A.W.B., Gruber,C., Jeesee,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished  Contact : Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: segre@genoscope.cns.fr Web : www.genoscope.cns.fr Library was constructed by life technologies, a division of invitrogen. This sequence belongs to sequene cluster 268.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/ccluster.cgi?seq=CSDBAKO54BD07NM1&cluster=268.f. Peng Liang Email : tiliang@lifeech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CSDBAKO54BD07NM1.			

FEATURES	source	Location/Qualifiers
		1..901
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="caxon:9606"
		/clone="CSOD1022YF01"
		/cissue_type="PLACENTA COT 25-NORMALIZED"
		/clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED"
		/note="1st strand cDNA was primed with a Noct-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT	326 a	139 c 123 g 309 t 4 others
ORIGIN		
Query Match	56.9%;	Score 479.4; DB 13; Length 901;
Best Local Similarity	70.2%;	Pred. No. 1.6e-78;
Matches	592; Conservative	0; Mismatches 245; Indels 6; Gaps 3
QY	5	AAAAAANTATTTAAATTTCTCTATTANCGTCTCCNCAANCATTATTNACCTATN 64
DB	7	AAAAAGCTATTTTAAATTTCTATTAAACATCTCTTAAAGCATTTATTTTCTCATA 66
QY	65	NONCNCANTTTNAAANTACCTTNNNTTAAAAAACCTNGSAAAAAATAATNGC 124
DB	67	TCTCCTAATTTTAAAGAAATACATAGATTAGAAAACTAGAAAAAGATTAATGC 128
QY	125	AAATANTTAACTTCTTNGAAAGAAATTTNTACCAANGACNGAAANCNTTAAAT 184
DB	127	AGATATTTAACTTTCATGAAAGAAATTTATATCAAGAAGCTGAGACGTTATAAT 188
QY	185	NGAANTAAATTAATANTTNGAAANGCGCNCNGAAACCAANTTTAATGCTCAATATCC 244
DB	187	TGAAATGAGATTAATTAATTTGAAACCTGCACTGAAAGCAAACTTTATGTTCAATATTC 248
QY	245	TNAAAGAGGANTTTNANNACTAATNCCNGATTTTCCATANGAANCNNNTTAAA 304
DB	247	TTAAAGATGGTGTTTATGACTTAATACACTGATTTTCAAGAAAGAAACCATGTTAAA 308
QY	305	NTNTTTTAAATTTTAAATAATTAACCCNGTNTCCAAACCCNATCANATTCCTTNAATTTGA 364
DB	307	ATATTTTATTTTAAATAATTAACCCCTGTGTTCAGAGCTGTATCATATTTCTTTATTTTGA 368
QY	365	TTGGGAAAAAATATNCNGTTCNNATACCCNNGAANNCAANTTTTAAATTTTAAACC 424
DB	367	TTTGGGAAGAAATATCTGTTTCTGATAGCATGAAATGCAAAATTTTATCTTTTAACT 428
QY	425	CCCTAATTTTAAATCTATNGAAATNTGATTAANAGCTTGAATTCG-CACCTTANTTN 484
DB	427	CACATAATTTTAAAGAACTATGGAATATGATTAATGACATGAGTCAACAACATAATTA 488
QY	484	CNGGCACACNGTGGGCTNTGTTTCTTACTTANTCCCCCAAGAAANNCTTAAACNG 544
DB	487	CTGGCAGAGCTGTGGCATGTGTCTTCTTACTTATGTTCTCCCAAGAAAACTCTTAAATG 548
QY	544	AANCTCCNCAAAATTAACCTTAATATCTTGGTAACCAANCAAAACCTTTTNGTTT 604
DB	547	AATCTTACGACAAATATCTTAAATATATCTTGTGAAGAAAAAACAAGCTTTTGTGTT 608
QY	604	ACNTAATCTTGGATTTAAGGGGTCCCAATTTNATTCNGAACCAANTTTTCCCCNAA 664
DB	607	ACATAGTCTTTGGANTTACTGTTCTTAATTTTATTCGAAACCTCAATTTTACCCGAGA 668
QY	664	CCATANTTACATTTTACCTTGTGAAGCAGCTGTTTGCANTNCCGAAANCAAGTANT 724
DB	667	CCATANTTACATTTTACCTTGTGAAGCAGCTGTTTGTATGCAATTTCCGACNAGCAGTAGT 728
QY	724	NTTCCC--CNGGCNCTTCCCCCGANCTTTGGGAAAAACGGGATNGGTCCGCCCC--TT 778
DB	727	ATACCATCAGCTCTATTCACCCAGACGATAGGAAAAACAGGATGTGTGACACCAATTTA 782
QY	779	AAAAAACAATTTTCCCTTTGGCCCAAGGNTTTTCCGCTCTAATTCGACATATA 838

Db 787 GAACAATCTATCCACGCTTTTGGCAATGGAAGTTTCAGTCAATTCGACATAA 846  
QY 839 AAA 841  
Db 847 AAA 849

RESULT 4  
LOCUS A1640146 821 bp mRNA linear EST 16-DEC-1999  
DEFINITION wa29f11.x1 NCI CGAP Kid1 Homo sapiens cDNA IMAGE:2299533 3' similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR (HUMAN); mRNA sequence.

ACCESSION A1640146  
VERSION A1640146.1 GI:4703255  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 821)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapdb-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: www.bio.lnlnl.gov/btrp/image/image.html  
Insert Length: 565 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 471.  
Location/Qualifiers  
1..821  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2299533"  
/lab\_host="DH10B"  
/clone\_lib="NCI-CCAP Kid1"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and as circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 132376-132391, 145607-145675, and 150552-150285). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 296 a 128 c 114 g 278 t 5 others  
ORIGIN

Query Match 56.7%; Score 477.6; DB 9; Length 821;  
Best Local Similarity 70.9%; Pred. No. 3.4e-78;  
Matches 581; Conservative 0; Mismatches 235; Indels 4; Gaps 3;

QY 2 GCCAAAAAANTTATTTTCTATTTCCTATTCCTCCGCAAAACATATTATTCCT 61  
Db 1 GCCAGAAAAAGTATTTTCTATTAACATCTCTCGAAGCATATTATTCCT 60  
QY 62 ATNNCNCNGANTTTNANAANTTACCTTNTTNTTAAAAACCTNGAAAAAATAT 121  
Db 61 AATATCTCACTGATTTTAAAGAAATACATTAATTTAGAAAAAGAAAAAGATAA 120  
QY 122 NCGAATATTTACCTTCTGAAAAAGAAATTTTACCAANGACNGAANCNTTNTA 181

Db 121 TGCAGATATTTAACTTACATGAAAAAGAAATTTATNAACAAGAGCTGAGAGCTTADA 180  
QY 182 ATTGCAANTTAAATATATNTTNGAANCGCNCNCAACCAANTTATGTGCAATTA 241  
Db 181 AATTGAAAAATGATTTATTAATTTGAAACCTGATCGAAAGCAAACTTTATGTTCAATTA 240  
QY 242 TCCCTAAGAGGNNTTTNNANCTAATNCCNGATTTTCCATFANGAANCCNNNTTA 301  
Db 241 TTCTTAATGATGCTGTTTATGACTATATACATGATTTTCAATAGAAACCATGTTA 300  
QY 302 AAANTNTTTTAAATTTAAAAAATAACCCNGTTCACCCCGATCANATTCCTTMAATT 361  
Db 301 AAAATATTTTATTTAAAAAATAAGCTGCTGTCAAGCTGATCATATTTCTTTAATT 360  
QY 362 GGATTGGGAAAAAATNCGNTCCNNATACCCNNGAANGCAANTTTTAAATTTTAA 421  
Db 361 TGATTGGGAAAAAATCTGTTTCTGTATACATGAAATGCAAAATTTTGAATTTTAA 420  
QY 422 CCCCCCTATTTTAAACATFNGAAANTGATTANNAGCTGATATGC-CAACCTTAN 480  
Db 421 TCTCACTAATTTTAAAGACTATTGAGAAATTTGATTAATGACATGAGTCACACACTTA 480  
QY 481 TTTCNGGCCACCGTGGCGCTNGTNTTCTTACTTANTCCCCCAAGGAANNCTTAN 540  
Db 481 TTACTGGCCAGCTGTGTGCATTTGTGTTCTTACTAGTTCGCCAAGAAAACTCTTAA 540  
QY 541 CNGAANTCTCNCCAAATATACCTTTAATATCTTGTGTACCAANCAAACTTTTNG 600  
Db 541 CTGAATCTTCAGCAAAATATATCTTAAATATCTTTGACCAAAACNAAAGCTTTTNG 600  
QY 601 TTTACNTA-NTCCTGGATTTAAGGGTCCCAATTTATTCNCAACCCANTTTTCCC 659  
Db 601 TTTACATAGTCTTTTGGATTTTACTGTCTTAATTTATCTGAATCTCAATTTACC 660  
QY 660 CNAACCATANTTACCATTTTACCTTGATAGCNCAGTGTGTCANTCCGCAANCAG 719  
Db 661 CAGACCATATATTACATATATACNTTTGTATGACAGTGTATGCAATTCGCCAAGCAG 720  
QY 720 TANTTTTC--CNGGNCCTTTCCCGGACNCTTGGAAGAAAAAGGATNGTCCCCCT 777  
Db 721 TAGATATACATCAGCTCTATTTCAACCCAGACATATAGAAAAAGATGGGCCACACAT 780  
QY 778 TAAAAAACACCTTCCCGCCTTTGGCCGAGNNTTNT 817  
Db 781 TAGAAAAATATCTATCCACGCTATTTGCCAAGTGTA 820

RESULT 5  
LOCUS BM985376 756 bp mRNA linear EST 20-FEB-2003  
DEFINITION UI-CF-EC1-acg-p-23-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone  
UI-CF-EC1-acg-p-23-0-UI 3', mRNA sequence.  
ACCESSION BM985376  
VERSION BM985376.1 GI:19611803  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 756)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548  
COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.researchgen.com) or from Open Biosystems  
 (www.openbiosystems.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 316-342, >AT rich#low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLY-A=yes.

# FEATURES

source

Location/Qualifiers  
 1..756  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-EC1-acg-p-23-0-UI"  
 /tissue\_type="lung"  
 /dev\_stage="Adult and Fetal"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-CF-EC1"  
 /note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-EC1 is a normalized cDNA library containing the  
 following tissue(s): Normal lung from adult and from fetal  
 day 64, day 87, week 19 and week 42. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pRT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 AAGTCTTAC.  
 TAG LIB=UI-CF-EC1  
 TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371  
 and 380-383  
 TAG\_SEQ=AAGTCTTAC"  
 BASE COUNT 268 a 107 c 97 g 283 t 1 others  
 ORIGIN

Query Match 56.0%; Score 472.2; DB 12; Length 756;  
 Best Local Similarity 73.2%; Pred. No. 3,4e-77;  
 Matches 533; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

2 GCCAATAAATTTTAAATTTCTATTAAACCTCCGCAACATTTATTACCT 61  
 19 GCCAATAAATTTTAAATTTCTATTAAACATTTCTTCAAGCATTTATTATCT 78  
 62 ATNNCNCNGAATTTTAAATTAACCTTNTNTTAAACCTTNGAATAAATAT 121  
 79 ATATCTCATGATTTTAAAGAAATTAACATTTAGTATTAGAAAAAAGATATA 138  
 122 NGCAATATNTTAACTTCTTGAAGAAATTTTATCAAGACGAAACCTTTTA 181  
 139 TGCAGATATATTAACCTTCAATGAAGAAATTAATTAACAAAGACGAGAACCTTTTA 198  
 182 ATTNGAATNTAAATTAATNTTGAAGAACGAGCAGCAACCTTAAATTTTCAATTA 241  
 199 AATTGAAATGAGATTAATTAATTTGAAGAACTGCTCGAAGCAAACTTTATTTTCAATTA 258  
 242 TCTTAAAGAGGAGNTTAAATTAATTCCTGATTTTCAATTAAGAACCCNNNTTA 301  
 259 TTCTTAATGATGTGTTTATGATTAATTAATTAATTTTCAATTAAGAAACCATGTTA 318  
 302 AAAATNTTTTAAATTTTAAATTAACCCGCTTCCAAACCCGCAATCAATTTCTTTATTT 361  
 319 AAAATATTTTAAATTTTAAATTAACCCGCTTGTTCAGGCTGATATATTTCTTTATTT 378  
 382 GGATTTGGGAAAAAATATCNGTTCNNATACCCNNAAGCAAAATTTTAAATTTTAA 421

Db 379 TGATTGGAGAAAAATACCTGTTCTGATACATGAATGCAAAATTTTAGATTTTAA 438  
 Qy 422 CCCCCCTAATTTTAAATCTTNGAAAAATNGATTANNGACTTGAATTC-CAACCCAN 480  
 Db 439 TCTACTAATTTTAAAGAACTTAATTAAGAAATGATTAATGATGAAGCAACACATTA 498  
 Qy 481 TTTCNGGACCAACGCTGGGCTGCTGCTTACTTAATCCGCCAAGAAANNCTTAAN 540  
 Db 499 TTACTGGGACGCTGTTGGCATTTGCTTTCTTACTTACTTACTTCCCAAGAAACTTTAA 558  
 Qy 541 CNGAATCTCCCAAAATTAACCTTAAATTAATCTTGTGTACCAACCAAACTTTTNG 600  
 Db 559 CTGAATCTTACGAGAAATTAATCTTAAATATATATCTTGTAAACCAAAACAAAGCTTTTNG 618  
 Qy 601 TTTACATNTCTGTTGGGATTTAAAGGGTCCCAATTTATATCNGAACCAATTTTCCGCC 660  
 Db 619 TTTACATAGTCTTGTGGATTTTACTGTTCTTAAATTTTATGAAACCAATTTTACC 678  
 Qy 661 NAACCATTAATTAACATTTTACTTGTGTAAGGCGCAGTGTGTCANTCCCAACAGT 720  
 Db 679 AGACCAATTAATTAACATTTTACTTGTATGACACAGTGTATGCAATTCGCCAAGCAGT 738  
 Qy 721 AATNTTCC 728  
 Db 739 AGTATACC 746

RESULT 6  
 LOCUS BM997078  
 DEFINITION UI-H-ED0-axo-f-03-0-UI s1 NCI CGAP\_ED0 Homo sapiens cDNA clone  
 IMAGE:5831426 3', mRNA sequence.  
 ACCESSION BM997078  
 VERSION BM997078.1 GI:19721979  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 761)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 CONTACT Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Jose Mercende  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 The following repetitive elements were found in this cDNA  
 sequence: 316-342, >AT rich#low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLY-A=yes.

## FEATURES

source

Location/Qualifiers  
 1..761  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5831426"  
 /tissue\_type="Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP\_ED0"  
 /note="Organ: Left Pubic Bone; Vector: pRT73-Pac  
 (Pharmacia) with a modified polylinker; Site 1: EcoR I;  
 Site 2: Not I; NCI-CGAP\_ED0 is a cDNA library containing  
 the following tissue(s): Chondrosarcoma cell line C55. The  
 library was constructed according to Bonaldo, Lennon and  
 Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer

containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAAGCT.

TAG\_LIB=UI-H-ED0  
TAG\_TISSUE=chondrosarcoma  
TAG\_SEQ=CGTCAAGCT

BASE COUNT 269 a 108 c 98 g 285 t 1 others  
ORIGIN

Query Match 56.0%; Score 472.2; DB 12; Length 761;

Best Local Similarity 73.2%; Pred. No. 3.4e-77; Indels 1; Gaps 1;

Matches 533; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

```

OY 2 GCCAAAAAANTATTATTTTCTATTAAACNTCTCCGCAAAACATTATTATTCCT 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19 GCCAGAAAAAGTATTTTATTTTCTATTAAACATTCTCTCAAGCATTATTTATCCT 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 62 ATNNCCNCGAANTTTNANAANTACCTTTNNNTTAAAAAAGCTNGAAAAAATAAT 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 79 ATATCTCACTGAATTTTAAAGAAATAACATTAGATTAGAAAACTGAGAAAAAGTAA 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 122 NCCAAATATTAACTTCTTGAAGAAATTTTACCAANGACNCAANCTTTTAA 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 TCCAGATTAATTAACCTTACATGAAAAAGAAATTTAACTGAGAACTGTA 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 182 ATTNGAANTAAATTAATNTTNGAAACGCGCNCNCAACCAACCTTATGTCATTA 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 199 AATTGAAAAGATTTATTTGAAACCTGATCTGAAAGCAACCTTTATTTGTCATTA 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 242 TCTTAAAGAGGANNTTNANNACTAATNCCNGATTTTCAATANGAANCCNNNTTA 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 259 TCTTAAATGATGTTGTTTATGACTAATACATGATTTTCAATAGAAACCCATGTA 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 302 AAANNTTTTAAATTAATTAACCCNGTNCACCAACCCNATCATNATCTTTNATTT 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 319 AAAATATTTTAAATTAATTAAGCTGTGTCAAGCTCTATATTTCTTTATTT 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 362 GATTGGGAAAAAATNCGNTTCCNNTATCCNNGAANNGCAAAATTTTAAATTTTAA 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 379 TGATTTGGGAGAAAAATCTGTTTCGATGACATGAAATGCAAAATTTTAAATTTTAA 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 422 CCCCCCTTATTTAAANCTATNGAANNTGATTAANGACTTGAATTC-CAACCTTAN 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 439 TCTCACTAATTTTAAAGAACTATTTGAATTTGATTAATGACATGAAGTGCACACACTTA 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 481 TTNCGGCCACCGNGGCGTNGTNTCTTACTTANCCCCCAAGGAAANNCCTTAN 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 499 TTACTGGCCAGCTGTGGCTGTGTTCTTACTTAACTTCCCAAGGAAAACTCTTAA 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 541 CNGAANCTCCNCAAAATTAACCTTAATATCTTGTGTAAACCAANCAAACTTTTNG 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 559 CTGAATCTTCACAGAAATAATCTTAAATATCTTTGTAGAGAAAAACAACCTTTTNG 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 601 TTTACNTANTCTTGGGATTTAAACGGGTCGCCAATTTNATCGNAAACCCANTTTTCCC 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 619 TTTACATAGTCTTGGGATTTTACTGTTCTTAATTTTATCTGAAACTCAATTTTACCC 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 661 NAACATANTTACCTTTTAACTGTTAGGCGNCGTNTTTCGANTCCGCAANCACT 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 679 AGACATATATTAACCATTTTACTTTGTATGACAGGTGTATGCAATTCGCAAGCAGT 738
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 721 AATNTTCC 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 739 ACTATACC 746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 7  
B0014192  
LOCUS B0014192 769 bp mRNA linear EST 26-MAR-2002

DEFINITION UI-H-ED1-axs-9-24-0-UI.s1 NCI CGAP\_ED1 Homo sapiens cDNA clone

IMAGE:5833007.3', mRNA sequence.

ACCESSION B0014192

VERSION B0014192.1 GI:19739093

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS 1 (bases 1 to 769)

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

COMMENT Unpublished

CONTACT: Robert Strausberg, Ph.D.

EMAIL: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

TISSUE Procurement: Dr. Jose Mercende

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

The following repetitive elements were found in this cDNA

Sequence: 317-343, >AT rich#Low\_complexity (matched complement)

Seq primer: M13 FORWARD

PolyA=yes.

FEATURES

source

Location/Qualifiers

1..769

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5833007"

/tissue\_type="Chondrosarcoma"

/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="NCI CGAP ED1"

/note="Organ: Left Pubic Bone; Vector: pT73-Pac

(Pharmacia) with a modified polylinker; Site 1: EcoR I;

Site 2: Not I; NCI CGAP\_ED1 is a normalized cDNA library

containing the following tissue(s): Chondrosarcoma cell

line C85. The library was constructed according to Bonaldo

, Lennon and Soares, Genome Research, 6:791-806, 1996.

First strand cDNA synthesis was primed with an oligo-dT

primer containing a Not I site. Double stranded cDNA was

ligated to an EcoR I adaptor, digested with Not I, and

cloned directionally into pT73-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is GCTCAAGCT.

TAG\_LIB=UI-H-ED1

TAG\_TISSUE=chondrosarcoma

TAG\_SEQ=CGTCAAGCT

BASE COUNT 272 a 110 c 100 g 285 t 2 others

ORIGIN

Query Match 55.5%; Score 468.2; DB 12; Length 769;

Best Local Similarity 72.7%; Pred. No. 1.8e-76;

Matches 543; Conservative 0; Mismatches 202; Indels 2; Gaps 2;

```

OY 1 GCCCAAAAAAANTATTATTTCTATTAAACNTCTCCGCAAAACATTATTATTCCT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19 GCCAGAAAAAGTATTTTATTTTCTATTAAACATTCTTCAAGCATTATTTATTC 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 61 TATNNCCNCGAATTTNANAANTACCTTTNNNTTAAAAAAGCTNGAAAAAATAAT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 79 TATATCTCACTGAATTTTAAAGAAATAACATTAGATTAGAAAAAAGTAA 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 121 TNGCAATATTTAACTTCTTGAAGAAATTTTACCAANGACNCAANCTTTT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 ATGCAATATTTAACTTACATGAAAAAGAAATTTTAACTGAGAACTGTAAT 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 181 AATTNGAANTAAATTAATNTTNGAAANCGCNCNCAACCAANCTTATGTCATTT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      199 AAATTGAATGAGATTATTAATTTGAAAATGATCGAAGAGAACTTATGTTCAATT 258
Qy      241 ATCTTGAAGAGAGGNNTTTANNACTTAATNCCNGATTTTCCAAATPANGAANCCNNNTT 300
Db      259 ATTCTTAATGATGGTGTATGACATAATACACTGATTTTTCATATPAGAAACCATGTT 318
Qy      301 AAAAATNTTTAATTTTAAATAAACCNGTTCGAACCCCGATCANTTCCCTTNAAT 360
Db      319 AAAAATATTTTAAATTTTAAATAAAGCCGTGTCTCAAGCTCGATCATATTTCTTTTAT 378
Qy      361 TGGATGGGGAATAAATNCGTTCNNATACCCNNGAANNGAAATTTTAAATTTTA 420
Db      379 TTGATTTGGGAAGAAATCTGTTCTGATAGAGAAATGCAAAATTTTATGATTTTA 438
Qy      421 ACCCCCTTAATTTTAAANCTATNGAATAATNGATTANNAGCTGAATGCG-CAACCTTA 479
Db      439 ATCTCACTAATTTTAAAGAACTATGAGAAATGATTAATGACATGAGACAAACACTA 498
Qy      480 NTTNCGGACACNGTGGGCGTNTTCTTACTTANCCCCCAAGAAANNCCTTA 539
Db      499 ATTACTGGCGACCTGTGGCATGTGTCTTACTTATGTTCTCCCAAGGAAACCTCTTA 558
Qy      540 NCGAANCTCCNCAAAATAACCTTAANTATCTTGTATACCAAAACAACTTTTNN 599
Db      559 ACTGAATCTTCAGCAAAATATCTTAATATATCTTGTATAGCAAAACAAAGCTTTT 618
Qy      600 GTTACNTANTCTCTTGGATTTAAGCGGTCCCAATTNATCCNGAACCCANTTTTCCC 659
Db      619 GTTACATATGTTCTTGGATTTTACTGTCTTAATTTTATCTGAAACCTCAATTTTACC 678
Qy      660 CNAACCATANTTACCTTTTACTTGTATAGCNCAGTGTGTGCANTGCCCAANCG 719
Db      679 CAGACCATATTAATCACTATTAACCTTGTATGACAGTGTATGCAAT-CCGANAAGCAG 737
Qy      720 TANTTTCCCGCNGCCTTCCCGCA 746
Db      738 TAGTATACATCAGGCTCTATTCACCA 764

RESULT 8
LOCUS   BX414370 905 bp mRNA linear EST 13-MAY-2003
DEFINITION BX414370 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
VERSION   CS0DM007YJ07 3-PRIME, mRNA sequence.
ACCESSION BX414370
KEYWORDS  BX414370.1 GI:30637003
SOURCE    EST.
ORGANISM  Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 905)
L1 W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 268.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAK077DD07NM1&cluster=268.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paradey Avenue Genoscope Sequence ID : CS0BAK077DD07NM1.
Location/Qualifiers
1..905
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

```

/clone="CS0DM007YJ07"
/ligase_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a Noci-oligo (dfr) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 333 a 136 c 123 g 312 t 1 others
ORIGIN
Query Match 55.3%; Score 466.2; DB 13; Length 905;
Beef Local Similarity 69.2%; Pred. No. 4.2e-76;
Matches 584; Conservative 0; Mismatches 252; Indels 8; Gaps 3;
8 AAAAATTTAATTTCTTAATTAACNTCTCCNCAANCAATTAATTTNACCTATNNGN 67
9 AAAAGAAATTCATATTTTAAATTAACATTAATCTCCAAAGCATTAATTTATCTATATCT 68
68 CNGCANNTTANNAANAATACCTTTNTTNTTAAAAAACCCTNGAAAAAATAATGCAAA 127
69 CACTGAATTTTAAAGAAATTAACATTAATTAAGAAAACTAGGAAAAAGATTAATGACA 128
128 TANTTAACCTTCTGAAANGAATTTTACAAAGACGAAACGTAATTAATTNGA 187
129 TAATTAACCTTACATGAAGAAAGAAATTAATACAAAGACTGAGAACCTTAATTAATGCA 188
188 ANTNAATTAATANTNGAANAACGCGNCGAAGAACCAACCTTATGATGCAATTAATCTTA 247
189 AATGATTAATTAATTTGAAAACTGCATCGAAAGCAACTTAATGTTCAATTAATCTTA 248
248 AAGAGGNNTTTANNACTAATNCCNGATTTTCCAAATPANGAANCCNNNTTAAATNTN 307
249 ATGATGGTGTATTAAGACTAATACATGATTTTCAATAAGGAAACCAATGTTAAAAAATA 308
308 TTTTAATTTTAAAAAATAACCCGNTTCCAAACCCGATCANATTCCTTATTTGGAATTG 367
309 TTTTATTTTAAAAATTAAGCTGTGTCTCAAGCTCGAACATATTTCTTTATTTTGATTT 368
368 GGGAAAAAATNCGTTCNNATACCCNNGAANNGCAANTTTTAAATTTTAAACCCCC 427
369 GGGAGAAAAATACGTCTTGTATGATACGAAATGCAAAATTTTATGATTTTAACTGAC 428
428 TANTTTTAAANCTATNGAATAATNGATTANNAGCTGAATGCG-CAACCTTAATNCG 486
429 TAATTTTAAAGAACTATGAGAAATGATTAATGACATGAAGTGCACACACTAATATCTG 488
487 GCCACGNTGGCGNNTTCTTACTTANCCCCCAAGAAANNCCTTAANCAGAN 546
489 GCCAGCTGTGGCATGTGTCTTACTTATGTTCTCCCAAGGAAACCTTTAACTGAAT 548
547 CTCNCCAAATAACCTTAAATATCTTGTGTAACCAAAACCTTTTNGTTTACN 606
549 CTTGACGAGAAATACCTTAATATATCTTGTATGACAAACAAAGCTTTTGTGTTTAA 608
607 TANTCTTGGATTTAAGCGGTCCCAATTNATCCNGAACCCANTTTTCCCCNAACCA 666
609 TAGTCTTGTGATTTTACTGTTCTTAATTTTATCTGAAACTTAATTTTAAACCCGACCA 668
667 TANTTAACATTTTACTTGTATAGCNCAGTGTGTCANTMCCCAANAGTANTNT 726
669 TAATTAACATTAATTAATTTGTAATGACAGNATGATGCAATTCGGAAGCAGTATATA 728
727 CCCCGG---CNCCTTCCCGCAGNCCTTGGGAAAAACGGATNGTCCCTCCCTTTAAAAA 783
729 CCATGACGCTCATTTATCCAGAGCATAGGAAAAACAGATTTGTGACGCCATTAAGAA 788
784 ACAACCTTCCCGCNC---TTGGCCCAAGNNTTNTTCCGCTTAATTCGAACATTA 839
789 ACAATATATACCAACCGCATATTTGCCAATGTGTAGTTTAAAGTCAATTAATTCGACATAAA 848
840 AAAG 843

```

Db 849 AATG 852

RESULT 9  
LOCUS A1826287  
DEFINITION wk35f07.x1 NCI CGAP Pr22 Homo sapiens cDNA clone IMAGE:2417221.3' similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR (HUMAN); contains element MER28 repetitive element; mRNA sequence.

ACCESSION A1826287  
VERSION A1826287.1 GI:5446958  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 770)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
TITLE Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Insert Length: 1454 Std Error: 0.00  
Seq primer: -40UP from Glibco  
High quality sequence stop: 404.  
Location/Qualifiers  
1..770  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2417221"  
/sex="male"  
/tissue\_type="normal prostate"  
/lab\_host="DH10B"  
/clone\_lib="NCI-CCAP\_Pr22"  
/note="Organ: prostate; Vector: pTT3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 271 a 111 c 99 g 285 t 4 others  
ORIGIN

Query Match 55.3%; Score 466; DB 9; Length 770;  
Best Local Similarity 72.5%; Pred. No. 4.7e-76;  
Matches 528; Conservative 0; Mismatches 199; Indels 1; Gaps 1;

2 GCCAATAAATATTTAATTCGATTAACNCTCCNCAANCAATTTATTTACCT 61  
|||||  
25 GCCAGAAAAGTTATTTATTTTCATTAACATTTCTCTAAGCATTTATTTATCT 84  
|||||  
62 ATNNCNCNCGANTTTNANAANTACCTTNNNTTAAAAAAGCTNCGAAAAAATAT 121  
|||||  
85 ATATCTCATGATTTTAAAGAAATACATTAGTATTTAGAAAAAGTAAAGATTA 144  
|||||  
122 NCGAAATATTTAATCTTCTTGAAGAAATTTTACCAANGACNCAANCTTTTA 181  
|||||  
145 TGCAGATATATTAATCATGATGAAAAAGAAATTTATTAACAAGCATGAGAAC 204  
|||||  
162 ATTGAATNNAATTAATNTTGAAGAACGCGNCGAAGCAACNCTTATGTGCAAT 241

Db 205 AATTGAATGAGATTATATTTGAAAACTGCATCGAAGCAAACTTATTTGTCATTA 264  
|||||  
242 TCTTNAAGAGGANNTTTANNACTAATNCCNGATTTTCAATNAGNCCNNNTTA 301  
|||||  
265 TTCTTAATGATGATTTTATGATGATATACATGATTTTTCATPAGGAAACCATG 324  
|||||  
302 AAATNTTTTATTTTAAAAATTAACCCNGTTCACCCCGATGATTTTCTTTNAT 361  
|||||  
325 AAAATATTTTATTTTAAAAATTAAGCCGTGTCAAGCTGATCATATTTCTTTAT 384  
|||||  
362 GGATGGGAAAAAATATNCGNTTCCNATPACNNGAANNCAANTTTTAAATTTAA 421  
|||||  
385 TGATTTGGGAAAGAAATATCTGTTCTGATGACATGAAATGCAAAATTTTATTT 444  
|||||  
422 CCCCCCTAATTTTAAANCTATNGAANAANTGATTANNAGCTGATTC-CAACCTTA 480  
|||||  
445 TCTCATATTTTAAAGACTATGGAATATGATTAATGACATGAAGTCAACACTAA 504  
|||||  
481 TTNCNGGCCACCGNGGCGNNTTCTTACTTANTCCCCCAAGAAANNCCTTA 540  
|||||  
505 TTACTGGCCACGTGTGGCATGTGTCTTACTTAAGTCTCCAAAGAAAACTTTAA 564  
|||||  
541 CNGAATCTCNCNCAAAATTAACCTTAATATCTTGTGAACCAANCAAACTTTTG 600  
|||||  
565 CTGAATCTTCAGCAGAAATATCTTTAATATATCTTTGTAAGCAAAACAAAGCT 624  
|||||  
601 TTACTNTATCTCTGGATTTAAGGGTCCCAATTNATCGNAAACCAATTTTCCC 660  
|||||  
625 TTATCATATGTTCTTTGATTTATGTTCTTAATTTATTTGAAACATCAATTTA 684  
|||||  
661 NAACATATTTACATTTTACTTGTGTGTAAGCNCAGNTTTGCANTNCCGAANCG 720  
|||||  
685 AGACCATATTAACCATATTAATCTGTATGACACAGTTGTATGCAATTTCCG 744  
|||||  
721 AATNTTCC 728  
|||||  
745 AGTATACC 752

RESULT 10  
LOCUS A1565988  
DEFINITION tn2a05.x1 NCI CGAP Ki411 Homo sapiens cDNA clone IMAGE:2171984.3' similar to gb:552450 NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR (HUMAN); mRNA sequence.

ACCESSION A1565988  
VERSION A1565988.1 GI:4524440  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 793)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
TITLE Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Insert Length: 654 Std Error: 0.00  
Seq primer: -40UP from Glibco  
High quality sequence stop: 458  
POLYA-No.  
Location/Qualifiers

source

1..793

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2171984"

/lab\_host="DH10B"

/clone\_lib="NCI CGAP Kid1"

/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and 96 circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 291 a 122 c 110 g 266 t 4 others

ORIGIN

Query Match 55.0%; Score 463.4; DB 9; Length 793;

Best Local Similarity 71.0%; Pred. No. 1.4e-75;

Matches 562; Conservative 0; Mismatches 224; Indels 5; Gaps 3;

1 GGGCAAAAAAATTTTAAATTTCTATTAACCTCCGCAACATTTATTTACCC 60

3 GGGCAAAAAAATTTTAAATTTCTATTAACCTCCGCAACATTTATTTACCC 62

61 TATNNCNCNGANTTTNNAANAATACCTTNNNTTAAACCTGGAACAAAAATTA 120

63 TATATCTCACTGAATTTTAAAGAAATACATTAAGTAAAGAAAGTAAAGTAA 122

121 TNGCAATATTTTAACTTCTGAAAGAAATTTTACCAAGACAGCAACATTTT 180

123 ATGCGATATATTAACCTGAGAAAGAAATTTTAAAGAGAGAGAGAGAGAT 182

181 AATTGAANTAAATTAATNTGGAAGGCGNCCGAAACCAACTTATGTCATTT 240

183 AATTGAATGAGATTAATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGATTT 242

241 ATCTTGAAGAGGNNTTNNAATTAATCCGATTTTCCATATGGAACCCNNNT 300

243 ATTCTTAATGAGGTTTATGACTAATACAGTATTTTCAATTAAGAAACCATGT 302

301 AAAAATTTTAAATTTTAAACCCGCTGTCACACCCGATCAATTCCTTTAAT 360

303 AAAAATTTTAAATTTTAAACCCGCTGTCACAGCTGATCATATTTCTTTAAT 362

361 TGGATTGGGAAAAAATNCGTCCNNATACNNGAANGAANTTTTAAATTTTA 420

363 TTGATTGGGAAAAAATACGTTTCTGATACATGAATGCAAAATTTTGAATTTTA 422

421 ACCCCCTAATTTTAAACATATGAAANTGATTAAGGACTGATGTC-CAACCTTA 479

423 ATCTCACTAATTTTAAAGATTAAGAAATGATTAAGATGAGGACCAACACTTA 482

480 NTTCNGGCGACCGCTGGGCTGTTCTTACTTATCCGCCCAAGAAANCCCTTA 539

483 ATTACTGGCGACGCTGTTGGCATTTGTTCTTACTTATCCGCCCAAGAAANCCCTTA 542

540 NONGAATCCGCAAAATTAACCTTAAATATCTTGGTAACCAAGAAANCCCTTTTN 599

543 ACTGAATCTTCAGCAAGAAATATCTTAATTAATCTTGGTAACCAAGAAANCCCTTTTG 602

600 GTTTAATTAATCTTCGGAATTAACGGGTCGCCCAATTTATCCGAAACCATTTTCCCC 659

603 G-TTACATAGCTCTTTGATTTAAGTGTCTTAATTTATCTGAAGCACTATTTAACC 661

660 CNAACCATATTTACATTTTACCTGGTAAGGCGAGTGTTCGCAANTGCCCAAGCAG 719

662 CAGACCATATTAATCACTTAATCTTGTATGACAGAGTGTATGCAATTTCCCAAGCAG 721

720 TANTTTTCCCGG---CNCCTTCCCGGAGCTTGGGAAAAACGGGATNGTCCCGCC 776

Db 722 TAGTATACATCAGGCTTATTCACCCAGACATTAAGAAAAACGAGTTGTCTACACC 781

Qy 777 TTAACCAACAA 787

Db 782 ATAGAAACAA 792

RESULT 11

LOCUS BUE27147

DEFINITION UI-H-FG0-bct-1-12-0-UI.s1 NCI CGAP EN1 2 Homo sapiens cDNA clone

UI-H-FG0-bct-1-12-0-UI 3', mRNA sequence.

ACCESSION BUE27147.1 GI:23293361

VERSION BUE27147.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 762)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)

Tissue Procurement: James Martin

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, [bento-soares@iowa.edu](mailto:bento-soares@iowa.edu)

The following repetitive elements were found in this cDNA sequence: 1-43, >AT rich#low complexity (matched complement) 309-335, >AT rich#low complexity (matched complement)

Seq primer: M13 FORWARD

POLY-A:yes

FEATURES

source

1..762

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-H-FG0-bct-1-12-0-UI"

/tissue\_type="Enchondroma cell line"

/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="NCI CGAP EN1 2"

/note="Organ: Bone; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP EN1 2 is a cDNA library containing the following tissue(s): Enchondroma cell line (2 cell lines). The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73D-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGGTCACTC. The cell lines was provided by Dr TAG Martin from University of Iowa.

TAG LIB=UI-H-FG0

TAG TISSUE=Enchondroma cell line (Mix of EN1 and EN2)

TAG SEQ=CGGTCACTC"

BASE COUNT 269 a 109 c 97 g 286 t 1 others

ORIGIN

Query Match 54.8%; Score 462.2; DB 13; Length 762;

Best Local Similarity 72.7%; Pred. No. 2.4e-75;

Matches 537; Conservative 0; Mismatches 200; Indels 2; Gaps 2;







/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-PH1-bfm-1-10-0-UI"  
 /tissue\_type="Cell Line"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI CGAP PH1"  
 /note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; NCI CGAP PH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)16 tail. The sequence tag for this library is AGAATCGGCG. The cell line was provided by Dr. James Martin from the University of Iowa.

AE1WORDS	ES1.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 822)  
Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z. and Han, Z.  
Homo sapiens CDNA HTC clones

TITLE  
JOURNAL  
COMMENT  
Unpublished  
Contact: Zenguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES  
source  
1..822  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HTCCAD08"  
/tissue\_type="Hypothalamus"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/clone\_1ib="HTC"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 292 a 133 c 115 g 277 t 5 others  
ORIGIN  
XhoI

Query Match 54.4%; Score 458.2; DB 9; Length 822;  
Best Local Similarity 71.4%; Pred.No.1.3e-74;  
Matches 552; Conservative 0; Mismatches 216; Indels 5; Gaps 3;

16 TTTNATTTCTATTAAACCTCCGCAAAACATATTATTTNACCCATNCCNCCGANT 75  
16 TTTTATTTTCTATTAAACATTTCTTCAAGACATTTATTTTCTCATATTCACGAT 75  
76 TTNANAANTACCTTTNTNTTAAACCTTNGAAGAAAAATTAATNGCAATATTAA 135  
76 TTTAAGAAATACATTAGATTAGAAAACTGAAAAAAGATAATGAGATTAATTA 135  
136 CTTCCTTGAAGAAATTTTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 195  
136 CTTCATATGAAAAAGAAATTTAACAAGGAGCTGAAAGCTTAATAAATTTGAATGAGAT 195  
136 TATATTTNANAAGCGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 255  
136 TATATTTGAAACATGATCTGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 255  
256 NTTNANNACTAATNCCGATTTTCAATANGAAGCAAGCAAGCAAGCAAGCAAG 315  
256 GTTTATATGACTAATACATGATTTTCAATAGAAAGCAAGCAAGCAAGCAAG 315  
316 TTTAAAAATTAAGCGCTGTGTTCAAGCTGTGATTTATTTATTTATTTGATTTGGAGAGA 375  
316 AATNCGCTTCCNNAATACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 435  
316 AATATCTGTTCTGATAGCATGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 435  
436 AAANCTATNGAAATNGATTANNGACTGAATTCG-CAACCTATNTNCCGACAGC 494  
436 AGAATCTATGAGAAATGATTAATGACATGAAGTGAAGCAAGCAAGCAAGCAAG 495  
436 TGGGCTGCTGTTCTTACTTANTCCCGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 554  
436 TTGGATGCTGTTCTTACTTANTCCCGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 555  
555 AATAATACCTTAATATCTTGTGTAACCAAGCAAGCAAGCAAGCAAGCAAGCAAG 614

Db 556 GATATATCTTTAAATATCTGTTAGACAAAACAAAGCTTTTGGTTACTATCTT 615  
Qy 615 GGGATTTAAGGGTCCCAATTTNATCCNAAACCAATTTTCCCAACCAATTTACC 674  
Db 616 TGGATTNATAGGT-CTTAATTTTATCTGAAATCAATTTAACCCAGACCAATTTACC 674  
Qy 675 ATTTACCTTGTAAGGCACTGTTTGCANTCCGCAAGCAAGCAAGCAAGCAAG 733  
Db 675 ATATTACCTGTAATGACAGATGATGACCTTCCGCAAGCAAGCAAGCAAGCAAG 734  
Qy 734 --CNCCTTCCCGCAAGCTTGGGAAAAACGGGATGTGCTCCCTTAAAAA 784  
Db 735 CTCTATTCACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 787

RESULT 15  
BQ045161 738 bp mRNA linear EST 21-FEB-2003  
LOCUS  
DEFINITION  
UI-CF-EN1-ae1-a-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone  
ACCESSION  
BQ045161  
VERSION  
BQ045161.1 GI:19796248  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 738)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL  
MEDLINE  
PUBMED  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548

COMMENT

Contact: McCray, PB  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@iowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).  
The following repetitive elements were found in this cDNA sequence: 316-342, >AT-rich#Low-complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=yes.

FEATURES  
source

1..738  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-ae1-a-17-0-UI"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_1ib="UI-CF-EN1"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not

1, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is CTGCTCAGT. TAG LIB=UI-CF-EN1

TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS 6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGT"

BASE COUNT 260 a 102 c 97 g 278 t 1 others  
ORIGIN

Query Match 54.0%; Score 455.6; DB 12; Length 738;  
Best Local Similarity 72.9%; Pred. No. 3.9e-74;  
Matches 517; Conservative 0; Mismatches 191; Indels 1; Gaps 1;

```
Qy 2 GCCAAAAAANTTATTNAATTTCTATTAAACCTCCNCAANCAATTATTNACCT 61
Db 19 GCCAGAAAAGTTATTTTCTATTAAACATTTCTCAAGCATTTATTTATCT 78
Qy 62 ATNNCNCNCNANTTTNAAAANTACCTTNNNTTAAAAAACCCTNGAAAAAATAT 121
Db 79 ATATCTCACTGAATTTTAAAGAAATTAACATTAATTAAGAAAACTAGAAAAAAGAT 138
Qy 122 NGCAATATNTTACCTTCTGAAAAAGAAATTTTACCAANGAGACNGAAANCNTT 181
Db 139 TGCAGATTAATTAACCTTACATGAAAAGAAAATTTATACAAAGACTGAGACCTTA 198
Qy 182 ATTNGAANTNAATTAATNTTNGAAANCGGNCNGAACCAANCTTATGATCCATTA 241
Db 199 AATTGAATGAGATTATATATTTGAAAACCTGACATGAAAAGCAAACTTATTTGTCATTA 258
Qy 242 TCCCTNAANGAGGNNTTNANNACTAATNCCNGATTTTCCATTAAGAAACCNNTTA 301
Db 259 TTCTTAAGATGGTGTTTTATGACTAATACATGATTTTTCATTAAGAAAACCATGTTA 318
Qy 302 AAATNTTTTATTTTAAAAAATACCCGNTTCCAAACCCNGATCANATTCCTTNAATTT 361
Db 319 AAAATATTTTATTTTAAAAAATAGCTGTGTTCAAGCTGATCATATTTCTTTATTT 378
Qy 362 GGATTGGGAAAAAATNCGNTTCCNATACCCNNGAANNGCAANTTTTAAATTTTAA 421
Db 379 TGATTTGGGAAGAAATACGTGTTCTGATAGCATGAATGCAAAATTTTGAATTTTAA 438
Qy 422 CCCCCCTTANTTTTAAANGCTATNGAAANTGATTANNGACTTGATTC-CAACCTAN 480
Db 439 TCTCACTAATTTTAAAGAACTATTGAGAAATGATTAATGACATGAAGTGCACAACACTTA 498
Qy 481 TTNCNGGCGACGNGGCGNTGTTCTCTACTTANTCCCCCAAGGAANNCCTTAAN 540
Db 499 TTACTGGCCAGCTGTGGCATGTGTGTTCTTACTTATGTTCTCCCAAGSAAAACCTTAA 558
Qy 541 CNGAANCTCNCNCAAAATACCTTAANTATCTTGTAAACCAANCAAAACCTTTTNG 600
Db 559 CTGAATCTTCAGCAATAATCTTAAATATACCTTGTGAGCAAAACAAAGCTTTTGG 618
Qy 601 TTTAACNTANTCTTGGGATTTAAAGGCTCCCAATTTNATCCNGAAACCAATTTTCCCCC 660
Db 619 TTTCATAGTCTTTGGATTTTACTGTCTTAATTTTATTCGAAACCTCAATTTTACCCC 678
Qy 661 NAACCATANTTACATTTTACCTGTGTAAGCNCAGTGTTCGANTNC 709
Db 679 AGACCATGATTAACCATTAATTAATTTGTAAGCAGTGTGATGCAATTC 727
```

Search completed: November 27, 2003, 12:28:20  
Job time : 2274.04 secs

**THIS PAGE BLANK (USPTO)**



PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
PI Pfirndechuh M, Sahin U, Scanlan MJ, Stockert E;  
PI Tureci O;  
XX WPI; 1999-132448/11.  
XX  
XX  
XX New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
XX  
XX  
XX Claim 67; Page 696; 787pp; English.  
XX  
XX The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
XX  
XX Sequence 843 BP; 237 A; 184 C; 89 G; 215 T; 118 other;

Query Match 86.0%; Score 725; DB 20; Length 843;  
Best Local Similarity 100.0%; Pred. No. 2,1e-162;  
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACAAAAAATTTTAAATTTCTTAATTAACCTCTCCCAAAACATTAATTNACCC 60  
DB 1 GGCACAAAAAATTTTAAATTTCTTAATTAACCTCTCCCAAAACATTAATTNACCC 60  
XX  
XX 61 TATNNCNCNCANNTTNAANAATACCTTNTNTTAAAAAACCTNGAAAAAATA 120  
DB 61 TATNNCNCNCANNTTNAANAATACCTTNTNTTAAAAAACCTNGAAAAAATA 120  
XX  
XX 121 TNGCAATATNTTAACTTCTGAAAGAAATTTTACCAANGAGACGAAGACCTTNT 180  
DB 121 TNGCAATATNTTAACTTCTGAAAGAAATTTTACCAANGAGACGAAGACCTTNT 180  
XX  
XX 181 AATTGAAATTAATTAATTTGAAAGCGCNCNGAACAACCAACTTATGTCATTT 240  
DB 181 AATTGAAATTAATTAATTTGAAAGCGCNCNGAACAACCAACTTATGTCATTT 240  
XX  
XX 241 ATCCCTAANAGAGGNNTTNANNACCTAATNCCNCAATTTTCCAAATANGAAGACCTTNT 300  
DB 241 ATCCCTAANAGAGGNNTTNANNACCTAATNCCNCAATTTTCCAAATANGAAGACCTTNT 300  
XX  
XX 301 AAAAATNTTTTAAATTAATAACCCGNTTCCAAACCCGATCANATTCCTTNAAT 360  
DB 301 AAAAATNTTTTAAATTAATAACCCGNTTCCAAACCCGATCANATTCCTTNAAT 360  
XX  
XX 361 TGGATTTGGGAAAAAATNCGTCCNATACCCNNGAANNGCAAAATTTTAAATTTT 420  
DB 361 TGGATTTGGGAAAAAATNCGTCCNATACCCNNGAANNGCAAAATTTTAAATTTT 420  
XX  
XX 421 ACCCCCTAATTTTAAANCTATNGAAGAAATGATTNAGACTTGAATGGCAACCTTAN 480  
DB 421 ACCCCCTAATTTTAAANCTATNGAAGAAATGATTNAGACTTGAATGGCAACCTTAN 480  
XX  
XX 481 TTNCNGGACACCGNGGCGTNGTNTTCTTACTTANTCCGCCAAGAAANCCCTTAN 540  
DB 481 TTNCNGGACACCGNGGCGTNGTNTTCTTACTTANTCCGCCAAGAAANCCCTTAN 540  
XX  
XX 541 CNGAANTCTCNCACAAATTAACCTTAATATCTTGGTAACCAAAACCAACCTTTTNG 600  
DB 541 CNGAANTCTCNCACAAATTAACCTTAATATCTTGGTAACCAAAACCAACCTTTTNG 600

QY 601 TTTACNTANTCCTTGGATTTAAGGGGTCCCAATTTTATNCCGAACCCANTTTTCCCC 660  
DB 601 TTTACNTANTCCTTGGATTTAAGGGGTCCCAATTTTATNCCGAACCCANTTTTCCCC 660  
XX  
XX 661 NAACCATANTTACATTTTACCTTGGTAAGGCNCAGTGTTCANTNCCGAANCACT 720  
DB 661 NAACCATANTTACATTTTACCTTGGTAAGGCNCAGTGTTCANTNCCGAANCACT 720  
XX  
XX 721 ANTNTTCCCGCNCNCCTTCCCGCAGAACCTTGGGAAAAACGGAGATNGTCCCCCTTAA 780  
DB 721 ANTNTTCCCGCNCNCCTTCCCGCAGAACCTTGGGAAAAACGGAGATNGTCCCCCTTAA 780  
XX  
XX 781 AAACACACCTTCCCGCNCNCCTTGGGCGCAGAGNNTTTCCGCTTAATCCGAACATATA 840  
DB 781 AAACACACCTTCCCGCNCNCCTTGGGCGCAGAGNNTTTCCGCTTAATCCGAACATATA 840  
XX  
XX 841 AAG 843  
DB 841 AAG 843

RESULT 2  
AAH57501/C  
ID AAH57501 standard; cDNA; 2663 BP.  
XX  
XX

AAH57501;

10-SEP-2001 (first entry)

Human liver cell specific cDNA sequence SEQ ID NO:341.

Human, tissue specific; diagnosis; brain; heart; skeletal muscle;  
lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;  
metabolic disease; developmental disease; cytostatic; immunomodulatory;  
neuroprotective; gene therapy; cancer; immunopathology; neuropathology.

Homo sapiens.

WO200132927-A2.

10-MAY-2001.

02-NOV-2000; 2000MO-US30396.

04-NOV-1999; 99US-0163508.

(INCY-) INCYTE GENOMICS INC.

Sornasse T, Selthamer J, Watson GA;

WPI; 2001-291057/30.

New cell and tissue specific polynucleotides useful for diagnosis,  
prognosis or monitoring of treatments for disorders where the gene is  
associated with a cancer, immunopathology or neuropathology -

Claim 1; Page 258-259; 327pp; English.

AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
sequences (I). (I) can have cytosolic, immunomodulatory and  
neuroprotective activities, and can be used in gene therapy. (I) and  
proteins (II) encoded by them are used in high throughput screening  
assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
mimetics, peptides, proteins, agonists, antagonists, antibodies or  
their fragments, immunoglobulins, inhibitors, drug compounds and  
pharmaceutical agents. Expression of (I) in a sample indicates the  
differentiation of embryonic stem cells into a tissue selected from  
brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
tissues. (I) and (II) are used to produce an expression profile that  
defines a metabolic or developmental process, treatment, condition,  
disease or disorder. The gene profile can be used for diagnosis,  
prognosis or monitoring of treatments and for investigating a  
predisposition to a disorder where the gene is associated with a

CC cancer, immunopathology or neuropathology.

**SQ** Sequence 2663 BP; 807 A; 472 C; 599 G; 785 T; 0 other;

Query match	Score	DB	Length
59.7%	503.4	22	2663

Best Local Similarity 71.7%; Pred. No. 9.5e-110;

Matches 609; Conservative 0; Mismatches 234; Indels 6; Gaps 3;

Oy	I	GGCCCAAAAAAANTTATTNNAAATTTGCTATTAANCNTCTCCGCACAANCAATTATTNNACC	60
Db	2653	GGCCCAAAAAAGTTATTTTTTAATTTTCATTTAAACATTTCTTCCAAGCATTTATTTATCTC	259
Oy	61	TATNCCNCNGCANFTTNAAAANAATACCTTNNNTTAAAAAACCTNGGAATAAAAAATPA	120
Db	2593	TATATCTACGTGAATTTTAGAAATPACCTTGATATAGAAAACCTAGAAAAAAGATPA	253
Oy	121	TNGCAATATNTTAACTTNCCTTGAAGAATAATTTNPACCAANGSACNGAANCNTNT	180
Db	2533	ATGCAGATAATTTAAACTTACATGAAAAAGAAAAATTATPACAAAGACTGAGAACGTTAT	247
Oy	181	AATTNGAANTNAATAATATNTNGAANAACCGCNNGCAAACCAACTNATGTGTCATTT	240
Db	2473	AAATTTGAATAGATTAATPATTTTGAATTAACCTGATCTGAAGGAACTTTATATGTTCAAT	241
Oy	241	ATCCNTAANGAGGANNTTNANNACTAATNCCNGATTTTCCATANGGAANCCNNNTT	300
Db	2413	ATTCTTAATGATGCTGCTTTTAGATCTAATACCTGATTTTTCAGAAAGAAACCATGTT	235
Oy	301	AAAAANTTTTNTATTTTAAAAATAACCCNGTNCACACCCNGATCANATTCCTTNAAT	360
Db	2353	AAAAATATTTTATTTTAAAAATAAGCTGTGTCAAGCTGTGATCATATTTCTTTTAT	229
Oy	361	TGGATGGGGAATAAATNCCNTCTCCNATACCCNNGANNGAANTTTTAAATTTTAA	420
Db	2293	TTGATTTGGGAAGAAATACGTGTTCTGATACATGAATGCAAAATTTTATGATTTTAA	223
Oy	421	ACCCCTCANTTTTAAANCTATNGAANAANTGATTAANGACTTGAATTC-CAACCTA	479
Db	2233	ATCTCACATATTTTAAAGACATTTAGAAATGATTAATGACATGAAAGCAACACATA	217
Oy	480	NITNCGGCACACNGTGGGCGNTGNTTCTTACTTANTCCCCCAAGAAANCCTTAA	539
Db	2173	ATTACTGGCCAGCTGTGGCATGTGTGTTCTTAATTAGTTCCTCAAGGAAAACTTPAA	211
Oy	540	NCNGAANTCCNCCAAATPAACCTTANATACCTTGGTAACCAACAAACCAATTTTNN	599
Db	2113	ACTGAATCTTACGACGAATPAATCTTAATAATACCTTGTGAACCAAAACAAAAGCTTTT	205
Oy	600	GTTTACNTANTCCTTGGGATTTAAAGGGTCCCACAAATTNATCCNGAACCCANTTTTCCC	659
Db	2053	GTTTACATAGTCTTTTGGATTTTACTGTTCCATATTTTATTTGTAACCTCAATTTTACC	199
Oy	660	CNAACCATANTTACCATTTTACCTTGTGAAGGCACAGTNGTTCGANTNCGCAANCG	719
Db	1993	CAGACCATATAATACATATTAACCTTGPATGACAGTGTATGCAATTCGCCAAAGCAG	193
Oy	720	TANTMTTCC--CCNGCNCCTTCCCCCGANCTTGGGAAAAACGGATNTGTCCTCCCT	777
Db	1933	TAGTATACCATCAGGCTCTATTACCCACAGAGATGAGAAAAACAGATTTGTACACCCA	187
Oy	778	T---AAAAAACMAACCTTCCCCCNCCTTTGGCCACAGANNNTNTTCCGTTAATCCGAAC	834
Db	1873	TTAAGAAACAATCTATCCCAAGCTATTTGCCAATGTGTAGTTTCAGTCTAAATCTGAC	181
Oy	835	AATAAAAAAG 843	
Db	1813	AATAAAATG 1805 .	

AC	AAK40094,
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Gastric cancer associated gene.
XX	
KM	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KX	prostate cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	MOJ904265-A2.
PD	
XX	
PF	28-JAN-1999.
XX	
PR	15-JUL-1998; 98WO-US14679.
XX	
PR	22-JUN-1998; 98US-0102322.
XX	
PR	17-JUL-1997; 97US-0896164.
XX	
PR	10-OCT-1997; 97US-0061599.
XX	
PR	10-OCT-1997; 97US-0061765.
XX	
PR	10-OCT-1997; 97US-0948705.
XX	
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDW-) LUDWIG INST CANCER RES.
P1	Chen Y, Gout I, Gare A, O'Hare M, Obata Y, Old LJ;
P1	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
P1	Tureci O;
DR	WPI, 1999-132448/11.
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
XX	for the diagnosis, monitoring or treatment of cancers
PS	Claim 67; Page 695; 787bp; English.
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
SQ	Sequence 714 BP; 258 A; 100 C; 92 G; 260 T; 4 other;
Query Match:	51.7%; Score 436.2; DB 20; Length 714;
Best Local Similarity	73.1%; Pred. No. 5.8e-94;
Matches 516; Conservative	0; Mismatches 187; Indels 3; Gaps 3
D5	2 GCCAATAAAATTTATTNAATTTTCTTAATNANCNTCCGCCCAANCAATTATTNACCCT 61
Y5	
D5	2 GCCGAGAAAAGTAGTATTTTAAATTTTCATATTAACAATTTCTTGCAAAGCATTTATTATCCT 61
Y5	
D5	62 ATNNCNCGNGANTTNANAANAANTAACCTTTNNNTTTAAAAAACCTNGGAAAAAATAAT 121
Y5	
D5	62 ATATCTCACTGAATTTTAAGAATAACACTTGTATTTAGAAAAACATGAGAAAAAGATAAA 121
Y5	
D5	122 NGCAATATNTTAACCTTNCCTTGAAANGAAATTTNTGCCANGACNGAANCAANTNTNA 161
Y5	
D5	122 TGCGATATATTAACATGAGAAAAAGAAAATATATTAACAAGAACTGAGAAACGTTATA 181
Y5	
D5	182 ATTNGAANTNAAATATATNTTNGAANGCGGCNCGAACCAGCAANTTNATGTGTCGAATA 241

```
Db 182 AATTGAAAAGATTTATTTGAAAACCTGCATCGAAGCAAACTTTTGTGCAATTA 241
Qy 242 TCCCTAANAGAGNNNTTANNACTAATNCCNGATTTTCCATANGAANCCNNNTTA 301
Db 242 TCCCTAATGATGGTGTATGACATTAATACATGATTTTCAAGAAAGCAACCATGTTA 301
Qy 302 AAAANNTTTTAAATTTAAATTAACCCNGTNTCCAAACCCCNATGCAATTCCTTATTT 361
Db 302 AAAATATTTTAAATTTAAATTAAGCTGTGTCAAGCTGTATCAATATTTCTTTATTT 361
Qy 362 GGAATGGGAAAAAATNCGTTCNNATACCNNGAANCAAAATTTTAAATTTTAA 421
Db 362 TGAATTTGGGAAAAAATTAATCTGTTTCGATAGCATGAAGCAAAATTTTAAATTTTAA 421
Qy 422 CCCCCCTAATTTTAAANCTATNGAANATTAANNGACTTGAATTCG-CAACCTTA 480
Db 422 TCTCACTAATTTTAAANACTATTTGAAGAAATGATTAATGACATGAAGTGCACAACTTA 481
Qy 481 TTNCNGCCACCGTGGGCTGNTGTTCTTACTTANTCCCCCA-GGAAANCTTAA 539
Db 482 TTACTGGCCAGCTGTGGCATGTGTTCTTACTTAAAGTCTCCCAAGGGAACCTTTAA 541
Qy 540 NCGAANCTCCNCAAAATTAACCTTAANTATCCTTGTGAACCAANCAAACTTTTAN 599
Db 542 ATTGAATCTTCAAGCAAAATTAATCTTAATATATCTTTGTAGCAAAACAAAGCTTTT 601
Qy 600 GTTTAATCTA-NTCTTTGGATTTTAAAGGCTCCCAATTTTATTCNGAACCCATTTTCC 658
Db 602 GTTTAATCTAATGTTCTTTGGATTTTACTGTTCCAAATTTTATTTCTGAACCTAATTTAC 661
Qy 658 CCNAACCAATTAATTAATCTTGTGTAAGGCAAGTGTTCG 704
Db 662 CCAAGCAATTAATTAATTAATTAATTTGTTTGTGCAAGTGTTCG 707

RESULT 4
AAK40095
ID AAK40095 standard; DNA; 687 BP.
XX
XX AAK40095;
AC
XX
XX 02-JUL-1999 (first entry)
DT
XX
XX Gastric cancer associated gene.
DE
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer; ss.
OS
XX Homo sapiens.
XX
XX WO9904265-A2.
XX
XX 28-JAN-1999.
PD
XX
XX 15-JUL-1998; 98WO-US14679.
PF
XX
XX 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obeta Y, Old LJ;
PI Pfeundschnh M, Sahin U, Scanlan MJ, Stockert E;
PI Tyreci O;
XX
XX WPI; 1999-132448/11.
```

```
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67; Page 695; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 687 BP; 242 A; 98 C; 85 G; 258 T; 4 other;
SQ
Query Match 48.8%; Score 411.8; DB 20; Length 687;
Best Local Similarity 71.8%; Pred. No. 3.5e-88;
Matches 495; Conservative 0; Mismatches 199; Indels 4; Gaps 3;
Qy 15 ATTNAATTTCCATTAANCTCTCCNCAANCAATTTTACCCTATNNCCNCGAN 74
Db 1 ATTNAATTTTCTAATTAACATTTCTTCAAGCATTTATTTATCCATCTCACTGAA 60
Qy 75 TTTNANAANACTCTTNTNTTAAACCTNGAAAAAATATATGCAATTTNTTAA 134
Db 61 TTTTANAANAATPAACATTAATGATTAAGAAAACTAGAAAAAGATNAATGCAATTAATTA 120
Qy 135 CTTTCTGAAAAANGAAATTTTACCAANGACAANGAANCNTTNTAATNGAANTNAAA 194
Db 121 ACTTACATGAAAAAGAAAAATTTATACAAAGACCTGAAACCTTATTAATGAAATGGA 180
Qy 195 TTAATNTTGAANCGGCGNCCNCAAGAACCAANCTTATGTCATATTCCTTAANGAGGG 254
Db 181 TTAATATTTTGAAGAACTGCATCGAAGCAAACTTATTTGTTCAATTTATCTTAATGATGG 240
Qy 255 NNTTANNACTAATNCCNCAATTTTCCATTAANGAANCCNNNTTAAATNTTAT 314
Db 241 TGTTTAAGACTAATPAACGATTTTTCATAAGGAAACCATGTTAAAAATATTTAT 300
Qy 315 TTTAAAAATTAACCGTNTTCCAAACCCNGATCANATTCCTTNAATTTGATTTGGGAAAA 374
Db 301 TTTAAAAATTAAGCCCTGTGTTCAGCTCGATCATATTTCTTTATTTTGAATTTGGGAAGA 360
Qy 375 AAATNCGTTCNNATACCNNGAANCAAAATTTTAAATTTTAAATCCCCCTANTTTT 434
Db 361 AAATACGTGTTCTGATAGCATGAAGTGAATTTTATGATTTTAAATCTCNCTAATTTT 420
Qy 435 AAAANCTATNGAANAATNGATTANNGACTTGAATTCG-CAACCTAANTNCGCCACCN 493
Db 421 AAGAACTATTAAGAAATTTGATTAATGACATGAAGTGAACCAACCTAATTTACTGGCGACT 480
Qy 494 GTGGGCTGNTGNTTCTTACTTANTCCGCCAAGAAANCCCTTAANCGAANCTCCGCC 553
Db 481 GTTGCAATGTGTGTTCTTACTTATGTTCTCCCAAGAAAACTTAAACTGAATCTTCAGC 540
Qy 554 AAAATTAACCTTAANTATCTTGTGAACCAANCAAAACCTTTTNGTTTACNTANTCT 613
Db 541 NGAATTAACCTTAATAATATCTTGTGAACCAAC--AAACCTTTTGTGTTACATAGTTCT 598
Qy 614 TGGATTTAAGGGTCCCAATTTNATCCNCAACCCANTTTTCCCCNCAACCAATNTAC 673
Db 599 TTGATTTTACTGTTCCATATTTATTTCTGAACCTCAATTTTCCCGACGACATTAATTAAC 658
Qy 674 CAT--TTTACCTTGGTAAGGCGAGTGT 701
```



Db 659 CCTATTAACTTTGTTATGACAGCTGT 687

RESULT 5

ABV87267  
ID ABV87267 standard; cDNA; 486 BP.

AC ABV87267;  
XX  
XX 13-DEC-2002 (first entry)  
XX  
XX Human colon cancer related cDNA SEQ ID NO 578.  
DE Human colon cancer related cDNA SEQ ID NO 578.  
XX  
XX Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;  
KM 88.  
XX  
XX Homo sapiens.  
OS  
XX  
XX MO200258534-A2.  
XX  
XX 01-AUG-2002.  
PD  
XX  
XX 19-NOV-2001; 2001WO-US43704.  
PF  
XX  
XX 20-NOV-2000; 2000US-252222P.  
PR 06-FEB-2001; 2001US-267011P.  
PR 28-MAR-2001; 2001US-279670P.  
PR 10-JUL-2001; 2001US-304037P.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX  
XX Stolk JA, Xu J, Chenault RA, Meagher MJ, Secretat H, King GE;  
XX  
XX WPI; 2002-608400/65.  
DR  
XX  
XX New isolated tumor colon polynucleotide and polypeptide, useful for the  
PT diagnosis, prevention and/or treatment of cancer, in particular colon  
PT cancer -  
XX  
XX Claim 1; SEQ ID NO 578; 266bp + Sequence Listing; English.

The invention relates to a human colon tumour expressed polynucleotide  
(I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of  
CC 2600 fully defined nucleotide sequences (ABV8665-ABV89289); (ii)  
CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)  
CC sequences that hybridize to (i), under moderately stringent conditions;  
CC (v) sequences having at least 75% or 90% identity to (i); or (vi)  
CC degenerate variants of (i). The compositions and methods of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC cancer, particularly colon cancer. (I) can be used in gene therapy and  
CC (I) and (II) are useful in pharmaceutical compositions such as vaccines.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 486 BP; 192 A; 55 C; 64 G; 175 T; 0 other;

Query Match 37.2%; Score 313.6; DB 24; Length 486;  
Best Local Similarity 72.8%; Pred. No. 6.1e-65;  
Matches 354; Conservative 0; Mismatches 131; Indels 1; Gaps 1;

QY 3 CCAAAAAAAAAATTTTAACTTTTCTTAAACNTCTCCNCAACCATTTTAAACCTTA 62  
XX  
XX  
XX  
Db 1 CCAAGAAAAAGTTATTTTATTTTCTTAAATATTTCTTCAAGCATTTATTTTCCCA 60  
XX  
XX  
XX  
QY 63 TNNCNCNGANTTTNANAANAATACCTTNNNTTAAAAACCTNGAATAAATAATN 122  
XX  
XX  
XX  
Db 61 TATTCACCTGAATTTTAAGAAATACATTAGATTGAAAACTAGAAAAAGATAAT 120  
XX  
XX  
XX  
QY 123 GCAAAATANTTAACCTTCTTGAAGAAATTTTACCAAGACGAAACNTTTTAA 182  
XX  
XX  
XX  
Db 121 GCAGATAATTAACCTTACATGAAAAAGAAATTTATACAAAGACTGAGACGTTATA 180

QY 183 TTNGAANTAAATATATANTNGAANAACGNNCNAACCAACCTTATGTCATAT 242  
XX  
XX  
XX  
Db 181 ATTGAATAGAGATTATATTTGAAACCTGATCGAAGCAACCTTTATGTCATAT 240  
XX  
XX  
XX  
QY 243 CCTNAAGAGGNNTTTANNACTAATNCCNGATTTTCAATANGAANCCNNNTTA 302  
XX  
XX  
XX  
Db 241 TCTTAATGATGATGTTTATGACTAATACACTGGTTTTCATATAGAAACCATGTTAA 300  
XX  
XX  
XX  
QY 303 AANTTTTNTATTTTAAAAATPACCCNGTNTCCAAACCCNGATCATMTTCTTTAATTTG 362  
XX  
XX  
XX  
Db 301 AATATTTTATTTTAAAAATAAGCCGTGTGCAAGCTGTGATCATATTTCTTTAATTTT 360  
XX  
XX  
XX  
QY 363 GATTGGGAAAAAATNCGTTCNNATACCCNNGAANNCAAAANTTTTAAATTTTAAAC 422  
XX  
XX  
XX  
Db 361 GATTGGGAGAAATTAAGTCTGTTTCTGATAGCATGAAATGCAAAATTTTATGATTTTAA 420  
XX  
XX  
XX  
QY 423 CCCCTANTTTTAAACATNAGAAANTNGATTANNGATTTGAATTCG-CAACCTANT 481  
XX  
XX  
XX  
Db 421 CTCCTAATTTTAAAGACTATTGAGAAATTTGATTATGACATGAGTGCACAACTAT 480  
XX  
XX  
XX  
QY 482 TNCNGG 487  
XX  
XX  
XX  
Db 481 TACTGG 486

RESULT 6

AAT19975/c  
ID AAT19975 standard; cDNA to mRNA; 351 BP.

AC AAT19975;  
XX  
XX  
XX  
DT 17-JUL-1996 (first entry)  
XX  
XX  
XX  
DE Human gene signature HMG501107.  
XX  
XX  
XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KM cell typing; abnormal cell function; 88.  
XX  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
XX  
PN WO9514772-A1.  
XX  
XX  
XX  
PD 01-JUN-1995.  
XX  
XX  
XX  
PF 11-NOV-1994; 94WO-JP01916.  
XX  
XX  
XX  
PR 12-NOV-1993; 93JP-0355504.  
XX  
XX  
XX  
PA (MATSU) MATSUBARA K.  
XX  
XX  
XX  
PI (OKUBO) OKUBO K.  
XX  
XX  
XX  
XX Matsubara K, Okubo K;  
XX  
XX  
XX  
DR WPI; 1995-206931/27.  
XX  
XX  
XX  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
XX  
XX  
PS Claim 1; Page 525; 2245pp; Japanese.  
XX  
XX  
XX  
XX A single-stranded DNA (or its complementary strand or the corresp.  
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences  
XX given in AAT19001-726837 and which is able to hybridize to part of  
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
XX sequences were obtained from 3'-directed cDNA libraries prepared  
XX from various human tissues; synthesis of cDNA was initiated from the  
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
XX untranslated sequence is unique to a particular mRNA species, almost  
XX all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library  
XX is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.

XX Sequence 351 BP; 117 A; 44 C; 44 G; 141 T; 5 other;

Query Match 25.6%; Score 216.2; DB 16; Length 351;  
 Best Local Similarity 73.6%; Pred. No. 6.8e-42;  
 Matches 256; Conservative 0; Mismatches 90; Indels 2; Gaps 2;

QY 1 GGGCAAAAAAANTATTATTTAATTCCTATTAAACCTCCCAAAACATTTATTNA-CC 59  
 DB 348 GGGCAAAAAAAGTTATTTTAACTTATTAACATTCCTCAAGCATTTATTATTTTANCC 289  
 QY 60 CTATNNCNCNGANTTTNANAANTACCTTTNNNTTAAAAAACCCTNGAAAAAATA 119  
 DB 288 CTATATCTCAGTAATTTTAAGAAATACATTAGTTTGAATAAAGTAAAGATA 229  
 QY 120 AATNGCAATANTTAACCTTCTTGAAGAAATTTTATCCAAAGACNGAAGNCNTTN 179  
 DB 228 AATGCGAATTAATTAACCTTACATGAAAAAATTATTAACAAAGACTGAAAGCTTA 169  
 QY 180 TAATTGAAANTNAATTTATANTTNGAAGAACGCGNCGAAGCAACCTTATGATGCTCAAT 239  
 DB 168 TAATTAATGAATGAGATTATTAATTTGAAGAACTGATCTGAAGCAACCTTATTTGTTCAAT 109  
 QY 240 TATCTTNAAGAGGNNNTTANNACTAAATCCGATTTTCCATATANGAAGAACCCNNNT 299  
 DB 108 TATCTTAATGATGCTTTTATGACTATACATGATTTTCAATTAAGAAACCATGT 49  
 QY 300 TAAANTNTTT--TATTTTAAAAATACCCGNTGCCAACCCCGATC 346  
 DB 48 TAAAAATATTTTNNNTTTAAAAATAAGCCTGTGTCAGCTGTGATC 1

## RESULT 7

AAH57293/c  
 ID AAH57293 standard; cDNA; 262 BP.

XX AAH57293;  
 AC XX  
 XX  
 DT 10-SEP-2001 (first entry)  
 XX  
 DE Human liver specific cDNA sequence SEQ ID NO:133.  
 XX  
 KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;  
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;  
 KW metabolic disease; developmental disease; cytostatic; immunomodulatory;  
 KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200132927-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 02-NOV-2000; 2000WO-US0396.  
 XX  
 PR 04-NOV-1999; 99US-0163508.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Sornasse T, Seilhamer J, Watson GA;  
 XX  
 DR WPI; 2001-291057/30.  
 XX  
 PT New cell and tissue specific polynucleotides useful for diagnosis,  
 PT prognosis or monitoring of treatments for disorders where the gene is  
 PT associated with a cancer, immunopathology or neuropathology -  
 XX  
 PS Claim 1; Page 117; 327pp; English.

XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
 CC sequences (I). (I) can have cytostatic, immunomodulatory and  
 CC neuroprotective activities, and can be used in gene therapy. (I) and  
 CC proteins (II) encoded by then are used in high throughput screening  
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
 CC mimetics, peptides, proteins, agonists, antagonists, antibodies or  
 CC their fragments, immunoglobulins, inhibitors, drug compounds and  
 CC pharmaceutical agents. Expression of (I) in a sample indicates the  
 CC differentiation of embryonic stem cells into a tissue selected from  
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
 CC tissues. (I) and (II) are used to produce an expression profile that  
 CC defines a metabolic or developmental process, treatment, condition,  
 CC disease or disorder. The gene profile can be used for diagnosis,  
 CC prognosis or monitoring of treatments and for investigating a  
 CC predisposition to a disorder where the gene is associated with a  
 CC cancer, immunopathology or neuropathology.

XX Sequence 262 BP; 103 A; 38 C; 34 G; 87 T; 0 other;

Query Match 20.1%; Score 169.6; DB 22; Length 262;  
 Best Local Similarity 74.7%; Pred. No. 7.1e-31;  
 Matches 195; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

QY 257 TTTNANNACTAATNCCGATTTTCCATFANGAAGAACCCNNNTTAAANTNTTNAATTT 316  
 DB 262 TTTTATGACTAATACACTGATTTTTCATTAAGAAACCATGTTAAATATTTTATTT 203  
 QY 317 TAAAAATTAACCCGNTTCCAAACCCGATCANATTTCTTTATTTGATGGGAAAAA 376  
 DB 202 TAAAAATTAAGCCTGTGTTCAAGCTGTGATCAATTTCTTTATTTGATGGGAAAAA 143  
 QY 377 ATNCGTCCNNATACCCNNGAANNCAANTTTTAAATTTTAAACCCCTANTTTTAA 436  
 DB 142 ATACGTTTCTGATGATGAAATGCAAAATTTTAAATTCACCTTAATTTTAA 83  
 QY 437 AANCTATNGAAANNGATTANNNGACTGAAATGTC-CAACCTANTTNGGCCACNGT 495  
 DB 82 GAACTATTGAGAAATGATTTATGACATGAAGTGCACAACTAATTCTGGCCAGCTGT 23  
 QY 496 GGGCNTNGNTTCTTACTTA 516  
 DB 22 TGGCAATGTGTTTCTTACTTA 2

## RESULT 8

ABX83093/c  
 ID ABX83093 standard; cDNA; 173 BP.

XX ABX83093;  
 AC XX  
 XX  
 DT 24-APR-2003 (first entry)  
 XX  
 DE Corn ear-derived polynucleotide (cpd) #1553.  
 XX  
 KW Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022;  
 KW SATMON023; structural gene; functional gene; regulatory gene;  
 KW corn ear-specific profile; gene transcription; gene expression;  
 KW hybrid plant; desirable trait expression; plant breeding program;  
 KW inheritance; desired characteristic; growth; development;  
 KW disease resistance; environmental adaptability; quality; yield;  
 KW multigene trait; plant; gene; ss.  
 XX  
 OS Zea mays.  
 XX  
 PN US6476212-B1.  
 XX  
 PD 05-NOV-2002.  
 XX  
 PF 14-MAY-1999; 99US-0313294.  
 XX  
 PR 26-MAY-1998; 98US-086722P.  
 XX



Db	187	ATTTCGCTATTTCCTACTACTTATTCTTAAACTCAATTTTACCCAAACCATGATTACCATA	246
Oy	678	TTACCTTGTAAGGCGCAGCTNGTTTGTCANTGCCAANAACAGTANTTTCC-CCNGGCNC	736
Db	247	TTAACTTTGTATATGACAGATTGTTTGCACCTCAGCAAGAGATGTAACCATCAGGTTTC	306
Oy	737	TTTCCCCCGCAGNCCTTGGGAAAAACGGGATNGTTCCTCCCTTAAAAAACACTTCCCC	796
Db	307	TATTACCCAGACACTTAGAAAAACGAATTGTATCACCCCATTTTAAAAACAATCTCAT	366
Oy	797	NCCTTGGCCGAGNNNTTNTCCCTCTAA	826
Db	367	GCTATTCCCAATGTAGAGTTCAATCTTA	396
RESULT 10			
ABX48095/C			
ID	ABX48095	standard; cDNA; 449 BP.	
AC	ABX48095;		
XX			
XX	21-FEB-2003	(first entry)	
DT			
XX			
DE	Bovine EST associated with lactation/muscle/fat deposition #13260.		
KW	Bovine; ss; EST; expressed sequence tag; lactation; IMF; muscle deposition; fat deposition; genome mapping; gene identification; gene analysis; cattle breeding.		
KW			
XX			
OS	Bos Taurus.		
XX			
PN	US2002137139-A1.		
PD	26-SEP-2002.		
XX			
PF	24-SEP-2001; 2001US-0960352.		
XX			
PR	12-JAN-1999; 99US-115707P.		
PR	11-JAN-2000; 2000US-0480902.		
XX			
PA	(BYAT/) BYATT J C.		
PA	(MATH/) MATHIALAGAN N.		
PA	(TAON/) TAO N.		
PA	(WARR/) WARREN W C.		
PI	Byatt JC, Mathialagan N, Tao N, Warren WC;		
XX			
DR	WPI; 2003-110599/10.		
XX			
PT	New nucleic acid associated with lactation, and muscle and fat deposition, useful for genome mapping, gene identification and analysis, cattle breeding, or for genetically improving cattle		
PT	analysis, cattle breeding, or for genetically improving cattle		
XX			
PS	Claim 2; SEQ ID No 13260; 245bp; English.		
XX			
CC	The invention relates to a purified nucleic acid molecule associated with lactation or muscle and fat deposition (designated IMF), derived from cattle, and the IMF nucleic acid can specifically hybridise to a second nucleic acid molecule comprising any of 15112 nucleotide sequences, appearing as ABX34836-ABX49947, or complements of them. Also included are; (1) a transformed cell having a nucleic acid comprising an IMF nucleic acid linked to a promoter and a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and (2) determining a level or pattern of a nucleic acid (comprising any of the 15112 nucleic acid sequences or its complement or fragment) with a complementary nucleic acid molecule obtained from the bovine cell or tissue, where hybridisation between the marker nucleic acid and the complementary nucleic acid permits the detection of the molecule; and (b) detecting the level or pattern of the complementary nucleic acid, where the detection of the complementary nucleic acid is predictive of the level or pattern of the molecule.		

CC	The LMD nucleic acid is used for determining a level or pattern
CC	of a molecule in a bovine cell or tissue. It is useful for genome
CC	mapping, gene identification and analysis, cattle breeding, preparation
CC	of constructs for use in cattle gene expression, or for genetically
CC	improving cattle. The present sequence is one of the 15112 bovine
CC	LMD EST (expressed sequence tag) nucleic acids.
CC	Note: The present sequence was not shown in the specification but
CC	was obtained in electronic format from the USPTO web site:
CC	seqdate.uspto.gov/sequence.html?DocID=20020137139.
XX	
SQ	Sequence 449 BP; 148 A; 70 C; 89 G; 142 T; 0 other;
Query Match	9.3%; Score 78.2; DB 25; Length 449;
Best Local Similarity	56.5%; Pred. No. 3.8e-09;
Matches 139; Conservative	0; Mismatches 106; Indels 1; Gaps 1
OY	TTNGTTCNTNANTCCTGGGATTTAACGGGTCGCCAATTNATCGNAACCAATTTTC 656
Db	430 TTCTGTTCACCATCTTTTGTAATTTTGTAACTTTATTTTAAACCTCAATTTTA 371
OY	657 CCCGNAACCATANTTTACCATTCTTGAAGGCNCAGTGTTCGANTNCCGCAAN 716
Db	370 CCCGACGACATGATTACATTTTAAGTGTATGACAGAGTTGTCAGACTCACAAAG 311
OY	717 CAGTANTTTC - CGNGCNCCTTCCCCGANCCTTGGAAAAACGGATNGTCCCCC 775
Db	310 CAGTAGTMAACCATCAGGTTCTATTCAACCAGACTTAGAAAACAGAAATTGATCACAC 251
OY	776 CTTAAAAACAACCTTCCCCCNCCCTTGGCCGAGGNNTTNTCCCGCTGAATCGGACA 835
Db	250 CCATTTAAAACAAACCTCATGCTATTCCCAGATGATAATGTTCAATCTAAGTCTGACA 191
OY	836 ATAAA 841
Db	190 ATGAAA 185
RESULT 11	
ID	ABX47851/C
XX	ABX47851 standard; cDNA; 397 BP.
AC	ABX47851;
XX	
DT	21-FEB-2003 (first entry)
DE	Bovine EST associated with lactation/muscle/fat deposition #13016.
KW	Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
XX	muscle deposition; fat deposition; genome mapping; gene identification;
OS	gene analysis; cattle breeding.
XX	
OS	Bos Taurus.
XX	
PN	US2002137139-A1.
XX	
PD	26-SEP-2002.
XX	
PF	24-SEP-2001; 2001US-0960352.
XX	
PR	12-JAN-1999; 99US-115707P.
PR	11-JAN-2000; 2000US-0480902.
PA	(BYAT/) BYATT J C.
PA	(MATH/) MATHIALAGAN N.
PA	(TAON/) TAO N.
PA	(WAR/) WARREN W C.
PI	Byatt JC, Mathialagan N, Tao N, Warren WC;
XX	
DR	WPI; 2003-110599/10.
XX	
XT	New nucleic acid associated with lactation, and muscle and fat
PT	deposition, useful for genome mapping, gene identification and

PT analysis, cattle breeding, or for genetically improving cattle -  
 XX  
 PS Claim 2; SEQ ID No 13016; 245bp; English.

CC The invention relates to a purified nucleic acid molecule associated with  
 CC laccation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridize to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX4836-ABX49847, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridization between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC [seqdata.uspto.gov/sequence.html?docID=20020137139](http://seqdata.uspto.gov/sequence.html?docID=20020137139).

CC  
 XX  
 SQ Sequence 397 BP; 124 A; 57 C; 80 G; 136 T; 0 other;

Query Match 8.3%; Score 69.6; DB 25; Length 397;  
 Best Local Similarity 56.2%; Pred. No. 4e-07;

Matches 163; Conservative 0; Mismatches 123; Indels 4; Gaps 3;

QY 553 CAAATTAACCTTAANTATCTTGGTAACCAANCAACCTTTTNGTTTACNTATGCC 612  
 DB 397 CAGAAATAGCTTCAATATACATTTATGCAAAATGAGAGCTTCT--GTTTACATACCTCA 340  
 QY 613 TTGGATTTAAGGGGCCCAATTNATCCGAACTTTTCCGCCCAACCAATANTTA 672  
 DB 339 TT-GAATTCGTATATTAATTAATCTTTTAAACCTTACCTTACCCCAACCAATGATTA 281  
 QY 673 CCATTTTACCTTGTAAAGCAGTNGTTGCANTNCCGCAANAGTANTTTCC-CCN 731  
 DB 280 CCATTTTAACTGTGTAAATGACAGATGTTAGCAACTGCAATGCTGTAAACATCA 221  
 QY 712 GGCNCTTTCCCGGAGCCTTGGGAAAAACGGGATNGTCCCTCCCTTAAAAACAACCTT 791  
 DB 220 GGTATATATTCACCCAGCAATTTGATTAACACAGCTTGATCACACCATTTAAAAACAAC 161  
 QY 792 CCCCNCCTTTGGCCGAGGNTTNTCCCGCTAAATCCGACCAATATAA 841  
 DB 160 CTCATGCTAATTTCCCAATCTATATGTTCAATCTAATCTTGAACAATGAAA 111

RESULT 12  
 AAQ76489/C  
 ID AAQ76489 standard; DNA; 263 BP.

XX  
 AC AAQ76489;  
 XX  
 DT 25-MAR-2003 (updated)  
 XX 23-SEP-1994 (first entry)  
 XX  
 DE Human genome fragment.  
 XX  
 KM Brain; placenta; bone marrow; genetic analysis; gene mapping;  
 XX detection; homology; human; adrenal tissue; de.

OS Homo sapiens.

XX  
 PN WO9401548-A2.

XX  
 PD 20-JAN-1994.

XX  
 PF 13-JUL-1993; 93WO-GB01467.

XX  
 PR 13-JUL-1992; 92GB-0014857.

XX  
 PA (MEDICAL RES COUNCIL.

XX  
 PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;  
 PI Sibson DR, Starkey M;

XX  
 DR WPI, 1994-035056/04.

PT New nucleic acid fragment encoding gene products - can be used  
 PT for genetic analysis and mapping

PS Claim 1; Page 64; 616bp; English.

CC Human nucleic acid fragments, isolated from brain adrenal tissue,  
 CC the placenta or bone marrow comprise any of: (A) a sequence  
 CC selected from (AAQ76401-077613), (B) an allelic variation of a  
 CC sequence as described in (A), or (C) a sequence complementary  
 CC to (A) or (B).  
 CC (Updated on 25-MAR-2003 to correct PN field.)

CC  
 XX  
 SQ Sequence 263 BP; 97 A; 38 C; 45 G; 80 T; 3 other;

Query Match 8.2%; Score 69.4; DB 15; Length 263;  
 Best Local Similarity 60.7%; Pred. No. 4e-07;

Matches 153; Conservative 0; Mismatches 92; Indels 7; Gaps 4;

QY 463 TTGAATGCCCAACCTTANTNCGGACACNGTGGCAGTNGTTCCTTACTTANTGCC 522  
 DB 263 TGGAGTACACGACGACTGATTTACTGGCCAGCTGTGGCATTTGTGTTTC-TACTTACTTTTC 205  
 QY 523 CCAAGAAANNCCTTAANCGAANCTCCNCCAAATAACCTTAANTATCTTGGTAAC 582  
 DB 204 ACGAGAAACCTTTAACTGAATCTTAGAGAAATCC--TTAATATCTTGTGA-C 149  
 QY 583 AAANCAAAACCTTTTNGTTTACNTANTCCTTGGATTTAACGGTCCCAATTNATCC 642  
 DB 148 AAAACAAAAGCTTTTGTGTTTACATCGTTCTTTGATTTTATGTTTATTTATTTCT 89  
 QY 643 NGAACCAANTTTTCCCGCAACCAANTTACCATTTTACCTTGTAGGAGCAGTGT 702  
 DB 88 GAATCTGATTTTACNNCCA--ACCATTAATTAATTTGTAATGACACAGTTGTAT 31  
 QY 703 GCANTNCCGCA 714  
 DB 30 GCAATTCAGAAA 19

RESULT 13  
 ABV60916  
 ID ABV60916 standard; cDNA; 340 BP.

XX  
 AC ABV60916;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 60907.  
 XX  
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 XX pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.



QY 185 NGAANTNAATTATANTNGAANGGNNCGAAGCAACTNATGTCATTAATCC 244  
 DB 16059 TAAATTAATAAATTCATAAACCATTTCATATATAACCAATTTTCACCTTACCAAAAAA 16000  
 QY 245 TNAANGAGGNNNTTANNACTAATNCCNGATTTTCCATANGAANCCNNNTTAAA 304  
 DB 15999 AATCTACATCTTACAAATTAATTAATTTTTCATTTTCAATTTTAAATGCAATAAAAA 15940  
 QY 305 NTNTTTNAATTTTAAATAATACCCGNTTCAACCCCGATCAATCTTCTTNAATTTGA 364  
 DB 15939 TCCAAATTAATCACACAAATTAACAAATTAATAAACAATCTATTTCTTTATTTTACA 15880  
 QY 365 TTGGGAAAAAATNCGTTCNNATACCNNAANGCAANTTTTAAATTTTAACCC 424  
 DB 15879 ATCATCTACTCTATCTATTTATTTTCAATCCATTAATTAATAAATACTAATTC 15820  
 QY 425 CCTTANTTTTAAAA 438  
 DB 15819 TACAAAAATTTTAAAA 15806

RESULT 15  
 ABL32941/C  
 ID ABL32941 standard; DNA; 18133 BP.

XX ABL32941;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 914.

XX Human, immune system disease; cytosine methylation; antiaesthetic;  
 KM antileukotectic; anti-HIV; anticonvulsant; ophthalmological;  
 KM antineoplastic; anti-HIV; anticonvulsant; ophthalmological;  
 KM antineoplastic; anti-HIV; anticonvulsant; ophthalmological;  
 KM antineoplastic; anti-HIV; anticonvulsant; ophthalmological;  
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KM gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -

PS Claim 1; SEQ ID NO 914; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

SQL Sequence 18133 BP; 5519 A; 331 C; 3750 G; 8533 T; 0 other;

Query Match 7.3%; Score 61.8; DB 24; Length 18133;

Best Local Similarity 40.1%; Pred. No. 7.3e-05;

Matches 174; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

QY 5 AAAAAAATTTTAAATTTTCTATTAATNCTTCGCAAGCAATTAATTTTAACTATN 64  
 DB 16239 AAAAAAATTTTAAATTTTCTATTAATNCTTCGCAAGCAATTAATTTTAACTATN 16180  
 QY 65 NCNCGNCAATTTTAAATTTTCTATTAATNCTTCGCAAGCAATTTTAACTATN 124  
 DB 16179 CACCTCTAATCAAAATCATTTTCAATATATCATATATATATATATATATTTT 16120  
 QY 125 AATATTTAATCTTCTGAAAGAAATTTTATCAAGCAAGCAAGCAAGCAAGCAAGCA 184  
 DB 16119 CTACTTTTATCTATTAATTAATAAATAATTTTAAATAATTTTCACTTTAAATA 16060  
 QY 185 NGAANTNAATTATANTNGAANGGNNCGAAGCAACTNATGTCATTAATCC 244  
 DB 16059 TAAATTAATAAATTCATAAACCATTTCATATATAACCAATTTTCACCTTACCAAAAAA 16000  
 QY 245 TNAANGAGGNNNTTANNACTAATNCCNGATTTTCCATANGAANCCNNNTTAAA 304  
 DB 15999 AATCTACATCTTACAAATTAATTAATTTTTCATTTTCAATTTTAAATGCAATAAAAA 15940  
 QY 305 NTNTTTNAATTTTAAATAATACCCGNTTCAACCCCGATCAATCTTCTTNAATTTGA 364  
 DB 15939 TCCAAATTAATCACACAAATTAACAAATTAATAAACAATCTATTTCTTTATTTTACA 15880  
 QY 365 TTGGGAAAAAATNCGTTCNNATACCNNAANGCAANTTTTAAATTTTAAACCC 424  
 DB 15879 ATCATCTACTCTATCTATTTATTTTCAATCCATTAATTAATAAATACTAATTC 15820  
 QY 425 CCTTANTTTTAAAA 438  
 DB 15819 TACAAAAATTTTAAAA 15806

Search completed: November 27, 2003, 06:23:52  
 Job time : 278.008 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2003, 06:06:02 : Search time 317.882 Seconds  
(without alignments)  
8737.270 Million cell updates/sec

Title: US-09-835-992A-22

Perfect score: 843  
Sequence: 1 ggcccaaaaantcattcttca.....taaatccgaacatacaaaag 843

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	725	86.0	843	9	US-09-835-992A-22
2	505	59.9	1052	10	US-09-880-107-3296
3	436.2	51.7	714	9	US-09-835-992A-19
4	411.8	48.8	687	9	US-09-835-992A-20
5	318.8	37.8	506	10	US-09-796-692-7228
6	318.8	37.8	506	14	US-10-040-862-7228
7	318.8	37.8	506	10	US-09-998-598-578
8	182.2	21.6	510	14	US-10-102-524-161
9	182.2	21.6	510	14	US-10-102-524-188
10	182.2	21.6	510	14	US-10-102-524-645
11	84.6	10.0	397	10	US-09-960-352-5962
12	78.2	9.3	449	10	US-09-960-352-13260
13	69.6	8.3	397	10	US-09-960-352-13016
14	61.8	7.3	18133	12	US-10-311-455-914
15	51.8	6.1	18218	12	US-10-311-455-1922
16	51.2	6.1	18434	12	US-10-311-455-1979

17	50.6	6.0	5452	12	US-10-311-455-1122	Sequence 1122, Ap
18	50.6	6.0	7351	12	US-10-311-455-2	Sequence 2, Appl
19	50	5.9	960	14	US-10-198-846-6381	Sequence 6381, Ap
20	49.6	5.9	9515	12	US-10-240-453-181	Sequence 181, Ap
21	49.6	5.9	9515	14	US-10-239-676-159	Sequence 159, Ap
22	49.4	5.9	9707	12	US-10-311-455-1393	Sequence 1393, Ap
23	49.2	5.8	3673778	12	US-10-312-841-2	Sequence 2, Appl
24	48.4	5.7	7201	12	US-10-311-455-309	Sequence 309, Ap
25	48.4	5.7	15667	12	US-10-311-455-2119	Sequence 2119, Ap
26	48.4	5.7	17491	12	US-10-240-485-128	Sequence 128, Ap
27	48.2	5.7	7049	12	US-10-311-455-130	Sequence 130, Ap
28	48.2	5.7	7049	12	US-10-240-452-6	Sequence 6, Appl
29	48	5.7	7781	12	US-10-311-455-1142	Sequence 1142, Ap
30	48	5.7	3673778	12	US-10-312-841-1	Sequence 1, Appl
31	47.4	5.6	5864	12	US-10-311-455-1413	Sequence 1413, Ap
32	47.4	5.6	5864	12	US-10-240-452-61	Sequence 61, Appl
33	47.2	5.6	6103	12	US-10-311-455-2143	Sequence 2143, Ap
34	47.2	5.6	18011	12	US-10-311-455-7	Sequence 7, Appl
35	47	5.6	6577	12	US-10-311-455-1329	Sequence 1329, Ap
36	47	5.6	6823	12	US-10-311-455-1118	Sequence 1118, Ap
37	47	5.6	10132	12	US-10-311-455-2053	Sequence 2053, Ap
38	47	5.6	10132	12	US-10-240-452-71	Sequence 71, Appl
39	46.8	5.6	6494	12	US-10-311-455-1366	Sequence 1366, Ap
40	46.8	5.6	10189	12	US-10-311-455-2133	Sequence 2133, Ap
41	46.6	5.5	10855	12	US-10-311-455-1461	Sequence 1461, Ap
42	46.2	5.5	884	10	US-09-070-927A-520	Sequence 520, Ap
43	46.2	5.5	6208	12	US-10-311-455-167	Sequence 167, Ap
44	46.2	5.5	9810	12	US-10-311-455-399	Sequence 399, Ap
45	46.2	5.5	19659	12	US-10-311-455-739	Sequence 739, Ap

#### ALIGNMENTS

RESULT 1  
US-09-835-992A-22  
Sequence 22, Application US/09835992A  
Patent No. US20020037541A1  
GENERAL INFORMATION:  
APPLICANT: Obata, Yutichi  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND  
TITLE OF INVENTION: METHODS FOR DIAGNOSING AND TREATING GASTRIC CANCER  
FILE REFERENCE: 10461/7112  
CURRENT APPLICATION NUMBER: US/09/835,992A  
CURRENT FILING DATE: 2001-04-16  
PRIOR APPLICATION NUMBER: US 08/896,164  
PRIOR FILING DATE: 1997-07-17  
NUMBER OF SEQ ID NOS: 87  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 22  
LENGTH: 843  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Unsure  
LOCATION: (12)..(12)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (19)..(19)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (33)..(33)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (35)..(35)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (42)..(42)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure  
LOCATION: (47)..(47)  
OTHER INFORMATION: n = a, c, g or t  
NAME/KEY: Unsure

```
LOCATION: (56) .. (56)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (64) .. (65)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (69) .. (69)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (71) .. (71)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (74) .. (74)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (78) .. (78)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (80) .. (80)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (84) .. (84)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (92) .. (93)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (95) .. (95)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (107) .. (107)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (122) .. (122)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (130) .. (130)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (139) .. (139)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (148) .. (148)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (157) .. (157)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (164) .. (164)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (169) .. (169)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (174) .. (174)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (176) .. (176)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (179) .. (179)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (185) .. (185)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (189) .. (189)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (191) .. (191)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (200) .. (200)
OTHER INFORMATION: n = a, c, g or t

NAME/KEY: Unsure
LOCATION: (203) .. (203)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (208) .. (208)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (213) .. (214)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (216) .. (216)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (225) .. (225)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (229) .. (229)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (246) .. (246)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (249) .. (249)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (255) .. (256)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (260) .. (260)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (262) .. (263)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (270) .. (270)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (274) .. (274)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (287) .. (287)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (292) .. (292)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (296) .. (298)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (305) .. (305)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (307) .. (307)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (312) .. (312)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (329) .. (329)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (332) .. (332)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (342) .. (342)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (348) .. (348)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (357) .. (357)
OTHER INFORMATION: n = a, c, g or t
```

```

/ NAME/KEY: Unsure
/ LOCATION: (379)..(379)
/ OTHER INFORMATION: n = a, c, g or t
/ NAME/KEY: Unsure
/ LOCATION: (381)..(381)
/ OTHER INFORMATION: n = a, c, g or t
/ NAME/KEY: Unsure
/ LOCATION: (387)..(388)
/ OTHER INFORMATION: n = a, c, g or t
/ NAME/KEY: Unsure
/ LOCATION: (394)..(395)
/ OTHER INFORMATION: n = a, c, g or t
/ NAME/KEY: Unsure
/ LOCATION: (399)..(400)
/ OTHER INFORMATION: n = a, c, g or t
/ NAME/KEY: Unsure
/ LOCATION: (406)..(406)

Query Match      86.0%; Score 725; DB 9; Length 843;
Best Local Similarity 100.0%; Pred. No. 2,7e-165;
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCACAAAAAATTAATTAATTCCTATTAAACCTCCNCAANCAATTAATTNACC 60
Db 1 GGCACAAAAAATTAATTAATTCCTATTAAACCTCCNCAANCAATTAATTNACC 60
Qy 61 TATNNCNCNGANTTTNANAANAATACCTTNTNTTAAAAAACCCTNGAAAAATPA 120
Db 61 TATNNCNCNGANTTTNANAANAATACCTTNTNTTAAAAAACCCTNGAAAAATPA 120
Qy 121 TNGCAATANTTAACCTTCTGAAAGAAATTTNTACCAAGACNGAAGAAACNTNT 180
Db 121 TNGCAATANTTAACCTTCTGAAAGAAATTTNTACCAAGACNGAAGAAACNTNT 180
Qy 181 AATTGAAANTAAATTAATTAATTTGAAACCGCNCNGAAACCAACCTTATGTCAT 240
Db 181 AATTGAAANTAAATTAATTAATTTGAAACCGCNCNGAAACCAACCTTATGTCAT 240
Qy 241 ATCTTNAANGAGGNTTTNANACTAATNCCGATTTTCCATTAANGAANCCNNNT 300
Db 241 ATCTTNAANGAGGNTTTNANACTAATNCCGATTTTCCATTAANGAANCCNNNT 300
Qy 301 AAAANTTTTNAATTTAAAAATAACCGNTTCCACCCCGATCANATTCCTTNA 360
Db 301 AAAANTTTTNAATTTAAAAATAACCGNTTCCACCCCGATCANATTCCTTNA 360
Qy 361 TGGATTGGGAAAAAATNCGTTCNNATAACNNGAANGCAANTTTTAAATTTTA 420
Db 361 TGGATTGGGAAAAAATNCGTTCNNATAACNNGAANGCAANTTTTAAATTTTA 420
Qy 421 ACCCCCTAATTTTAAANCTAATNGAANAANTGATTANNGACTGAATTCCTAN 480
Db 421 ACCCCCTAATTTTAAANCTAATNGAANAANTGATTANNGACTGAATTCCTAN 480
Qy 481 TTNCGGCGACCGGCGGNTGTTTCTTAATNCCCGCAAGAAANNCCTTAN 540
Db 481 TTNCGGCGACCGGCGGNTGTTTCTTAATNCCCGCAAGAAANNCCTTAN 540
Qy 541 CNGAANCTCNCACCAAAATACCTTAATATCTTGTAAACCAANCAAAACCTTTNG 600
Db 541 CNGAANCTCNCACCAAAATACCTTAATATCTTGTAAACCAANCAAAACCTTTNG 600
Qy 601 TTTACNTATCTTGGGATTTAAACGGGTCCCAATTTNATCCNGAACCCANTTTTCCC 660
Db 601 TTTACNTATCTTGGGATTTAAACGGGTCCCAATTTNATCCNGAACCCANTTTTCCC 660
Qy 661 TTTACNTATCTTGGGATTTAAACGGGTCCCAATTTNATCCNGAACCCANTTTTCCC 660
Db 661 TTTACNTATCTTGGGATTTAAACGGGTCCCAATTTNATCCNGAACCCANTTTTCCC 660
Qy 721 ANTNTTCCCGGCGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 780
Db 721 ANTNTTCCCGGCGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 780

```

```

Qy 781 AAAACAACCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Db 781 AAAACAACCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Qy 841 AAG 843
Db 841 AAG 843

RESULT 2
US-09-880-107-3296/c
Sequence 3296, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
PRIOR FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 3296
LENGTH: 1052
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U11313
US-09-880-107-3296

Query Match      59.9%; Score 505; DB 10; Length 1052;
Best Local Similarity 71.8%; Pred. No. 4.9e-112;
Matches 610; Conservative 0; Mismatches 233; Indels 6; Gaps 3;

Qy 1 GGCACAAAAAATTAATTAATTCCTATTAAACCTCCNCAANCAATTAATTNACC 60
Db 1028 GGCACAAAAAATTAATTAATTCCTATTAAACCTTCTCAAGCAATTAATTATCC 969
Qy 61 TATNNCNCNGANTTTNANAANAATACCTTNTNTTAAAAAACCCTNGAAAAATPA 120
Db 968 TATATCTCAGTAATTTAAAGAAATTAACATTAGTATGAAAAAAGTAAAGAAATPA 909
Qy 121 TNGCAATANTTAACCTTCTGAAAGAAATTTNTACCAAGACNGAAGAAACNTNT 180
Db 908 ATGAGATTAATTTAACTTACATGAAAAAGAAATTAATAAGAAAGACTGAGAACGTTAT 849
Qy 181 AATTGAAANTAAATTAATTAATTTGAAACCGCNCNGAAACCAACCTTATGTCAT 240
Db 848 AATTTGAATAGATTAATTAATTTGAAACCTCATCTGAAAGAAACCTTATTTGTCAT 789
Qy 241 ATCTTNAANGAGGNTTTNANACTAATNCCGATTTTCCATTAANGAANCCNNNT 300
Db 788 ATCTTATATGATGAGGTTTATGACTAATACATGATTTTCAATTAAGAAACCATGTT 729
Qy 301 AAAANTTTTNAATTTAAAAATAACCGNTTCCACCCCGATCANATTCCTTNA 360
Db 361 AAAANTTTTNAATTTAAAAATAACCGNTTCCACCCCGATCANATTCCTTNA 360
Qy 361 TGGATTGGGAAAAAATNCGTTCNNATAACNNGAANGCAANTTTTAAATTTTA 420
Db 668 TGGATTGGGAAAAAATNCGTTCNNATAACNNGAANGCAANTTTTAAATTTTA 609
Qy 668 TGGATTGGGAAAAAATNCGTTCNNATAACNNGAANGCAANTTTTAAATTTTA 609
Db 668 TGGATTGGGAAAAAATNCGTTCNNATAACNNGAANGCAANTTTTAAATTTTA 609
Qy 421 ACCCCCTAATTTTAAANCTAATNGAANAANTGATTANNGACTGAATTCCTAN 479
Db 608 ATCTTCAATTAATTTAAAGAACTATGAGAAATTTGATTAATGACATGAGAACATCA 549

```

Qy	480	NTNNNGGCGCCACCGNGGGCTNGTNGTTCCTTACTTANTCCCGCCCAAGAAANNCTTAA	539
Db	548	ATTACTGGCCAGCTGTGGCATGTGTCTTACTTAGTCTCCCAAGAAAACCTTTAA	488
Qy	540	NCNGAANCCTCCNCCAAATPACCTTAAATCCTTGTAAACAAACAAACCTTTTN	599
Db	488	ACTGAATCTTCAGCAGATATATCTTAATATATCTTGTAAACAAACAAACGTTTTT	429
Qy	600	GTTTACTTANTCCTTGGGATTTAAGGGTCCCAATTATTCNGAACCCANTTTTCCC	658
Db	428	GTTTACATATCTTTTGGATTATTAACGTCTCCAAATTTATCTGAACTCAATTTTACC	368
Qy	660	CNAACCATANTTACATTTTACCTTGTAAGGNCAGTNGTTGCANTCCGCAAAACAG	719
Db	368	CAGGCAATATTAACCATATTAATCTTTGTATGCAAGTTGTATGCAATTCGCAAAAGCG	309
Qy	720	TANTNTTCC--CCNGGNCCTTCCCGCCAGNCCTTGGGAAAAAAGGAGATNGTCCCGCT	777
Db	308	TAGATATACCATCAGGCTCTATTACCCAGAGATAGAAAAAAGAGATTGTCACACCA	249
Qy	778	T-----AAAAACAACCTTCCCCCNCCTTGGCCAGAGANTTNTTCCGCTTAATCGAAC	834
Db	248	TTAAGAAACAATCTATCCAGCGCTATTGCCGATGTGTGATTCAGCTAAATTCGAC	189
Qy	835	AATPAAAAAG	843
Db	188	AATPAAAAATG	180

```

RESULT 3
US-09-835-992A-19
Sequence 19, Application US/09835992A
Patent No. US20020037541A1
GENERAL INFORMATION:
APPLICANT: Obata, Yuichi
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND
FILE REFERENCE: L0461/7112
CURRENT APPLICATION NUMBER: US/09/835,992A
PRIOR APPLICATION NUMBER: US 08/896,164
PRIOR FILING DATE: 1997-07-17
NUMBER OF SEQ ID NOS: 87
SOFTWARE: PatentIn version 3.0
SEQ ID NO 19
LENGTH: 714
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Unsure
LOCATION: (243)..(243)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (373)..(373)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (437)..(437)
OTHER INFORMATION: n = a, c, g or t
NAME/KEY: Unsure
LOCATION: (702)..(702)
OTHER INFORMATION: n = a, c, g or t
US-09-835-992A-19

Query Match      51.7%; Score 436.2; DB 9; Length 714;
Best Local Similarity 73.1%; Pred. No. 1.7e-95;
Matches 516; Conservative 0; Mismatches 187; Indels 3; Gaps 3;

OY      2  GCCAAAAAATTTTATTTCCTATTANCNCTCCGCAACANCAATTTTACCCT 61
Db      2  GCCAGAAAATTTTATTTCCTATTTAACAATCTCTCCAAAGCATTTTATTCCT 61
OY      62  ATNNCNCNCANTTTNANAANAATDCTTTNNNTTTAAAAAACCTNGAAAAAAATAT 121

```

D6	62	ATATCTCAGCAATTTTAAAGAAATAACATTAGTATTAAGAAAACTAGAGAAAAAAGATPAA	121
Oy	122	NGCAATANTTTAACTTCTTNGAANAANGAAATTTNTACGANGACNGAANNTNTA	181
D6	122	TGAGATTAATTAACCTTACATGAAAAAGAAAAATTTATACAAGAGACTGAGAACGTATA	181
Oy	182	ATTNGAANTNAAATTAATANTTNGAANGGCGNCGNCGAACCANACTNTATGTCCTAATTA	241
D6	182	AATTGAAAATGAGATTATTAATTTGAAAACTCGATCTGAAGAGAACTTATTTGTCATTA	241
Oy	242	TCCTNANGAGGAGNNTTANNACTAATNCCNGATTTTCCAATANGAANCCNNNTA	301
D6	242	TNCTTATATGATGGTGTGTTTATGACTAATAACATGATTTTTCAAGAAAGAAACCATGTA	301
Oy	302	AAANTTTTNTATTTTAAAAATAACCGNTGCCAACCCNGATCANATTCCTTNTATTT	361
D6	302	AAATATTTTTTTATTTTAAAAATMACCGTGCTCAAGCTCGATCATAATTTCTTTATTTT	361
Oy	362	GGAATGGGGAATAAAATNCNGTTCNNATPACCNNGAANGCAANTTTTTTAAATTTTAA	421
D6	362	TGATTTGGGAAAAAATACTGTTTCTGATGACGAAAATGCAAAATTTTATGATTTTAA	421
Oy	422	CCCCCCTANTTTTAAANCTATNGAAANNTGATTAANGACTGAATTC-CAACCTAN	480
D6	422	TCTCACTAATTTTAAANAACTATGTAGAAATGATTAATGATGATGACACAACATTA	481
Oy	481	TTCNCGGCCACGNGGCGTNGTNGTTCCTACTANTCCCCCAA-GGAANNCCCTTA	539
D6	482	TTACTGGCAGCTGTGGCATTTGTCTTACTTAATGTTCTCCCAAGGAAAACTCTTAA	541
Oy	540	NCGAACTCCNCCAAATATACCTTAANTATCCCTGGTAAACAAACAAACCTTTTN	599
D6	542	ATTGAATCTTCAGCGAATAATCCTTAATATCTTTGTAAGCAAAACAAAGCTTTTTT	601
Oy	600	GTTTACNTA-NTCTTGGGATTTAAGSGTCCCAATTTNATCCNGAACCCANTTTTTCC	658
D6	602	GTTTACATAGTCTCTGGGATTTTACTGTCTCAATTTATTTCAAGAACTCAATTTTACC	661
Oy	659	CNAACCATANTTACATTTTACCTTGTGAAGGCGNGCTGTTGC-704	
D6	662	CCAGACCATAATTAACATTAATCTTGTTNGACAGGTGTGTTGC-707	

```

1 RESULT 4
2 US-09-835-992A-20
3 Sequence 20, Application US/09835992A
4 Patent No. US20020037541A1
5 GENERAL INFORMATION:
6 APPLICANT: Obata, Yuichi
7 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH GASTRIC CANCER AND
8 TITLE OF INVENTION: METHODS FOR DIAGNOSING AND TREATING GASTRIC CANCER
9 FILE REFERENCE: 10461/7112
10 CURRENT APPLICATION NUMBER: US/09/835,992A
11 CURRENT FILING DATE: 2001-04-16
12 PRIOR APPLICATION NUMBER: US 08/896,164
13 PRIOR FILING DATE: 1997-07-17
14 NUMBER OF SEQ ID NOS: 87
15 SOFTWARE: PatentIn version 3.0
16 SEQ ID NO 20
17 LENGTH: 687
18
19 TYPE: DNA
20
21 ORGANISM: Homo sapiens
22
23 FEATURE:
24 NAME/KEY: Unsure
25 LOCATION: (67)..(67)
26 OTHER INFORMATION: n = a, c, g or t
27
28 NAME/KEY: Unsure
29 LOCATION: (105)..(105)
30 OTHER INFORMATION: n = a, c, g or t
31
32 NAME/KEY: Unsure
33 LOCATION: (412)..(412)
34 OTHER INFORMATION: n = a, c, g or t
35
36 NAME/KEY: Unsure

```

LOCATION: (541)..  
OTHER INFORMATION: n = a, c, g or t  
US-09-835-992A-20

Query Match 48.8%; Score 411.8; DB 9; Length 687;  
Best Local Similarity 71.8%; Pred. No. 1.4e-89;  
Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;

```
QY 15 ATTAAATTTCTATTAAATCCTCCNCAANCAATTTATTTAACTTATNNCNCNCAN 74
DB 1 ATTTAAATTTCTATTAAATCCTCCNCAANCAATTTATTTAACTTATNNCNCNCAN 60
QY 75 TTTNAAATTAACCTTTNNNTTAAATTAATATATATATATATATATATATATAT 134
DB 61 TTTTAAATTAACCTTTNNNTTAAATTAATATATATATATATATATATATATAT 120
QY 135 CTTTCTGAAATGAAATTTTATCAAGACGAAATTTTAAATTTAAATTTAAAT 194
DB 121 ACTTACATGAAATGAAATTTTATCAAGACGAAATTTTAAATTTAAATTTAAAT 180
QY 195 TTATATNTTGAATGAAATGAAATTTTATCAAGACGAAATTTTAAATTTAAAT 254
DB 181 TTATATTTTGAATGAAATGAAATTTTATCAAGACGAAATTTTAAATTTAAAT 240
QY 255 NNTTANNAATTAATNCCNGATTTTCCATATGAAATCCNNTTAAATNTTTAT 314
DB 241 TGTTTATGATCTTAATACCTGATTTTCTATATAGAAATCCATTTAAATATTTTAT 300
QY 315 TTTTAAATTAACCTTNGTTCACACCCNGATCANATTCCTTATTTGATTTGGGAAA 374
DB 301 TTTTAAATTAACCTTNGTTCACACCCNGATCATATTTCTTTATTTGATTTGGGAAA 360
QY 375 AAATGCTGCTCCNATACCCNCAANCAATTTTAAATTTTAAATTTTAAATTTT 434
DB 361 AAATGCTGCTCCNATACCCNCAANCAATTTTAAATTTTAAATTTTAAATTTT 420
QY 435 AAATGCTGCTCCNATACCCNCAANCAATTTTAAATTTTAAATTTTAAATTTT 493
DB 421 AAATGCTGCTCCNATACCCNCAANCAATTTTAAATTTTAAATTTTAAATTTT 480
QY 494 GTGGCCTGCTGCTTCTTATTTTCCCAAGAAATTTTAAATTTTAAATTTTAAAT 553
DB 481 GTGGCCTGCTGCTTCTTATTTTCCCAAGAAATTTTAAATTTTAAATTTTAAAT 540
QY 554 AAATGCTGCTCCNATACCCNCAANCAATTTTAAATTTTAAATTTTAAATTTT 613
DB 541 AAATGCTGCTCCNATACCCNCAANCAATTTTAAATTTTAAATTTTAAATTTT 598
QY 614 TGGGATTTAAGGCTCCCAATTTTATCCNCAATTTTCCCAAGAAATTTTAAAT 673
DB 599 TGGGATTTAAGGCTCCCAATTTTATCCNCAATTTTCCCAAGAAATTTTAAAT 658
QY 674 CAT-TTTACCTTGTAAAGCAGTNGTT 701
DB 659 CCTATTAACTTTGTTATGCAAGTTGT 687
```

RESULT 5  
US-09-796-692-7228  
Sequence 7228, Application US/0979692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Manion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
PRIOR FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479

PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 7228  
LENGTH: 506  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-7228

Query Match 37.8%; Score 318.8; DB 10; Length 506;  
Best Local Similarity 73.4%; Pred. No. 3.6e-67;  
Matches 358; Conservative 0; Mismatches 129; Indels 1; Gaps 1;

```
QY 1 GGCAGAAAAATTTATTTAAATTTCTTATTAATCCTCCNCAANCAATTTTAAAT 60
DB 19 GGCAGAAAAATTTATTTAAATTTCTTATTAATCCTCCNCAANCAATTTTAAAT 78
QY 61 TATNNCNCNCATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAAT 120
DB 79 TATATCTCCTGAAATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTT 138
QY 121 TNGCAATTAATTAATCCTTCTGAAATGAAATTTTAAATTTTAAATTTTAAATTT 180
DB 139 ATGCAATTAATTAATCCTTCTGAAATGAAATTTTAAATTTTAAATTTTAAATTT 198
QY 181 AATGAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 240
DB 199 AATGAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 258
QY 241 ATCCNCAAGGCTTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 300
DB 259 ATCTTAATGATGCTTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 318
QY 301 AAAATNTTTTAAATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 360
DB 319 AAAATNTTTTAAATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 378
QY 361 TGGATTTGGAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 420
DB 379 TGGATTTGGAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 438
QY 421 ACCCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 479
DB 439 ATCTCAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 498
QY 480 NTNNNGG 487
DB 499 ATTTCTGG 506
```

RESULT 6  
US-10-040-862-7228  
Sequence 7228, Application US/10040862

```

/ Publication No. US20030078396A1
/ GENERAL INFORMATION:
/ APPLICANT: Gaiger, Alexander
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Mannion, Jane
/ APPLICANT: Retter, Marc
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
/ TITLE OF INVENTION: Hematological Malignancies
/ FILE REFERENCE: 014058-015200S
/ CURRENT FILING DATE: 2001-11-06
/ PRIOR APPLICATION NUMBER: US 60/186,126
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: US 60/190,479
/ PRIOR FILING DATE: 2000-03-17
/ PRIOR APPLICATION NUMBER: US 60/200,545
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: US 60/200,303
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,779
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,999
/ PRIOR FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: US 60/202,084
/ PRIOR FILING DATE: 2000-05-04
/ PRIOR APPLICATION NUMBER: US 60/206,201
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: US 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: US 60/222,903
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: US 60/223,416
/ PRIOR FILING DATE: 2000-08-04
/ PRIOR APPLICATION NUMBER: US 60/223,378
/ PRIOR FILING DATE: 2000-08-07
/ PRIOR APPLICATION NUMBER: US 09/796,692
/ PRIOR FILING DATE: 2001-03-01
/ NUMBER OF SEQ ID NOS: 10467
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 7228
/ LENGTH: 506
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-040-862-7228

```

Query Match	37.8%	Score 318.8	DB 14	Length 506
Best Local Similarity	73.4%	Pred. No. 3,66-67		
Matches 358	Conservative	0	Mismatches 129	Indels 1
Qy	1	GGCCAAAAAAANTTATTTTAAATTTCTCTATTAANCNTCCCTCCCAANCAATTAATTTNACC	60	
Db	19	GGCCGAGAAAAAGTTATTTTAAATTTCTATTAACAATCTTCTCCAAAGCATTTATTTATCC	78	
Qy	61	TATNNCCNCNGANTTTNAAAAAANTACCTTNNNTTAAAAAACCCTNGAAAAAAATTA	120	
Db	79	TATATCTCACTGAATTTTAAAGAAATAACCTGTAGTTAGAAAAACATGAGAAAAAAGATTA	138	
Qy	121	TNGCAATANTTAACCTTNCCTTGAAGAAANGAAATTTNTACAAAGACNGAAGAAANTNT	180	
Db	139	ATGCAGATPAATTTAACTTACATGAAAAAGAAAAATTAATACAAAGACTGAGAAAGCTTAT	198	
Qy	181	AATTGAAANTYAAATTTATANTTNGAAGAACGGCNCNGAAACCAACCTTAAATGTCGAAT	240	
Db	199	AAATTTGAATGAGATTATPATTTATTTGAAAACTGCATCTGAAGCAAACTTATTTGTCAAAT	258	
Qy	241	ATCTTNAAGAGGANNTTTANNACTAAATNCCGCAATTTTCCAATANGAAGAACCCNNNT	300	
Db	259	ATTCCTTAATGAGTGCTTTATGACTAATACACTGATTTTTCATTAAGAAAACCATGTT	318	
Qy	301	AAANNTNTTTTAAATTTTAAAAATTAACCCNGNTTCCAAACCCCGATCAAAATTCCTTNAAT	360	
Db	319	AAAAATATTTTATTTTAAAAATAAGCCTGTGCTTAAAGCTCGATCAATATTTCTTTAT	378	

Qy	361								
Db	379	TTGATTTGGAGAAATAACTGTCTTCTGATGCATGAATGCAAAAATTTTTAGATTTTA	438						
Qy	421	ACCCCCCTANTTTTAAAAAANCTAINGMAAANTGATTANNGACTTGCC-CAAACCTTA	479						
Db	439	ATTCACATTAATTTTAAGAACAATAATTSAGAAATTGATTAAATGACATGAAGTGCACAAACACTA	498						
Qy	480	NTTNCGG	487						
Db	499	ATTACTGG	506						
 RESULT 7 US-09-998-598-578 ; Sequence 578, Application US/09998598 ; Patent No. US20020150922A1 ; GENERAL INFORMATION: ; APPLICANT: Stolk, John A. ; APPLICANT: Xu, Jiangchun ; APPLICANT: Chenault, Ruth A. ; APPLICANT: Meagher, Madelein Joy ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND ; FILE REFERENCE: 210121.561 ; CURRENT APPLICATION NUMBER: US/09/998,598 ; CURRENT FILING DATE: 2001-11-16 ; NUMBER OF SEQ ID NOS: 2606 ; SOFTWARE: Corixa Invention Disclosure Database ; SEQ ID NO 578 ; LENGTH: 486 ; TYPE: DNA ; ORGANISM: Homo sapiens US-09-998-598-578									
Query Match	37.2%	Score 313.6;	DB 10;	Length 486;					
Best Local Similarity	72.8%;	Pred. No. 6.3e-66;							
Matches 354; Conservative	0;	Mismatches 131;	Indels 1;	Gaps 1					

Query Match	37.2%	Score 313.6	DB 10	Length 466
Best Local Similarity	72.8%	Pred. No. 6.3e-6		
Matches 354	Conservative	0	Mismatches 131	Indels 1
Qy	3	CCAAAAAANNTATTTNNAATTTCCATTAATNAACNTCCCTCCNCAAAACATTAATTTNACCTGA	62	
Db	1	CCAGAAAAAGTATTTTAAATTTTCATTAATAATCTTCTCAAGCATTAATTTTATCCGA	60	
Qy	63	TNNCNCNCNGANTTTNANAAAANTACCTTNNNTTTAAAAAAACCTNGAAAAAAAATAATN	122	
Db	61	TATCTCACTGAATTTTAAAGAAATAACATTAGATATAGAAAAACTAGAAAAAAGATAAAT	120	
Qy	123	GCAATATTTAACCTTCTTGTAAAAANGAAATTTMTACCAANGACNCAAAACNTTNTTA	182	
Db	121	GCGAGTAATTTAACTTACATGAAAAAGAAAAATTATTAACAAAGACTAGAACGTTATTA	180	
Qy	183	TTNGAANTNAATTAATANTTNGAANGCGCNCNCAAAACCAACCTTNAATGTCCAATTAT	242	
Db	181	ATTGAAATGAGATTTATTAATTTGAAAACCTGCATCTGAAGCAAACTTTATTTGTCATTAAT	240	
Qy	243	CCYNAANGAGGNNTTTANNANCTAATNCCNGAATTTTCCAATNGAANCCCNNTTAA	302	
Db	241	TCTTAATGATGTTGTTTATGACTATACACTGCGTTTTCAATAGAAGAAACCATGTTAA	300	
Qy	303	AAANTTTTNTAATTTAAAAATPAAACCCGNTTCCAAACCCCGATCANATTTCTTTNATTTG	362	
Db	301	AAATATTTTATTTTAAAAATPAAAGCGTGTCTCAAGCTCTGATCATATTTCTTTATTTT	360	
Qy	363	GATTTGGGAAAAAATNCGNTTCCNNATACCNNGAANGCAAAATTTTAAATTTTAAAC	422	
Db	361	GATTTGGGAAAAAATACCTGTGTTCTGAATAGATGAATGCAAAATTTTAAATTTTAAAT	420	
Qy	423	CCCCCTAATTTTAAAAANCATNGAANAANTGATTANNACCTGAATTC-CAACCCATNT	481	
Db	421	CTCACTAATTTTAAAGAACTATTGAGAAATTAATTTATATGACATGAAGTGCACAACTAAT	480	
Qy	482	TNCGNG	487	

Db 481 TACTGG 486

## RESULT 8

US-10-102-524-161/c  
; Sequence 161, Application US/10102524  
; Publication No. US20030109434A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Gordon, Brian  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER  
; FILE REFERENCE: 210121.572  
; CURRENT APPLICATION NUMBER: US/10/102,524  
; CURRENT FILING DATE: 2002-03-19  
; NUMBER OF SEQ ID NOS: 1863  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 161  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-102-524-161

Query Match 21.6%; Score 182.2; DB 14; Length 510;  
Best Local Similarity 69.5%; Pred. No. 3.9e-34;  
Matches 248; Conservative 0; Mismatches 104; Indels 5; Gaps 2;

Oy 492 CNGTGGGCGTNGTNTCTTCTTAATTCCTCCCAAGAAANNCTTAANGAANCCTCN 551  
Db 510 CTGTGGGATGTGTCTTCTTAATTCCTCCCAAGAAANNCTTAATGAATCTTCA 451  
Oy 552 CCAAAATACCTTAATATCTTGTGTAACCAAAACAACTTTTNGTTTACTANTC 611  
Db 450 GCAAAATATCTTAAATATCTTGTGTAACCAAAACAACTTTTNGTTTACTANTC 391  
Oy 612 CTGTGGGATTTAAGGGTCCCAATTTATTCGAAACCAATTTTCCCAACATANTT 671  
Db 390 CTGTGGGATTTAAGGGTCCCAATTTATTCGAAACCAATTTTCCCAACATANTT 331  
Oy 672 ACCATTTTACCTTGTGTAAGGCGAGTGTGTAACCAATTTTCCCAACATANTTCC--C 729  
Db 330 ACCATTTTACCTTGTGTAAGGCGAGTGTGTAACCAATTTTCCCAACATANTTCC 271  
Oy 730 CNGGCGCTTTCCCGGAGTGTGTAACCAATTTTCCCGTCTAATTCGAACATTAATAAG 786  
Db 270 AGGCTCTATTCACCGAGATGTAAGCAATTTGTCACACCCATTAAGAAACAA 211  
Oy 787 ACCTTCCCGGAGTGTGTAACCAATTTTCCCGTCTAATTCGAACATTAATAAG 843  
Db 210 TCTATCCACGCTATTTGCGCAATGTGTAGTTCAGTCTAATTTGACAAATTAATG 154

## RESULT 9

US-10-102-524-188/c  
; Sequence 188, Application US/10102524  
; Publication No. US20030109434A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Gordon, Brian  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER  
; FILE REFERENCE: 210121.572  
; CURRENT APPLICATION NUMBER: US/10/102,524  
; CURRENT FILING DATE: 2002-03-19  
; NUMBER OF SEQ ID NOS: 1863  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 188  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-102-524-188

Query Match 21.6%; Score 182.2; DB 14; Length 510;  
Best Local Similarity 69.5%; Pred. No. 3.9e-34;  
Matches 248; Conservative 0; Mismatches 104; Indels 5; Gaps 2;

Oy 492 CNGTGGGCGTNGTNTCTTCTTAATTCCTCCCAAGAAANNCTTAANGAANCCTCN 551  
Db 510 CTGTGGGATGTGTCTTCTTAATTCCTCCCAAGAAANNCTTAATGAATCTTCA 451  
Oy 552 CCAAAATACCTTAATATCTTGTGTAACCAAAACAACTTTTNGTTTACTANTC 611  
Db 450 GCAAAATATCTTAAATATCTTGTGTAACCAAAACAACTTTTNGTTTACTANTC 391  
Oy 612 CTGTGGGATTTAAGGGTCCCAATTTATTCGAAACCAATTTTCCCAACATANTT 671  
Db 390 CTGTGGGATTTAAGGGTCCCAATTTATTCGAAACCAATTTTCCCAACATANTT 331  
Oy 672 ACCATTTTACCTTGTGTAAGGCGAGTGTGTAACCAATTTTCCCAACATANTTCC--C 729  
Db 330 ACCATTTTACCTTGTGTAAGGCGAGTGTGTAACCAATTTTCCCAACATANTTCC 271  
Oy 730 CNGGCGCTTTCCCGGAGTGTGTAACCAATTTTCCCGTCTAATTCGAACATTAATAAG 786  
Db 270 AGGCTCTATTCACCGAGATGTAAGCAATTTGTCACACCCATTAAGAAACAA 211  
Oy 787 ACCTTCCCGGAGTGTGTAACCAATTTTCCCGTCTAATTCGAACATTAATAAG 843  
Db 210 TCTATCCACGCTATTTGCGCAATGTGTAGTTCAGTCTAATTTGACAAATTAATG 154

## RESULT 10

US-10-102-524-645  
; Sequence 645, Application US/10102524  
; Publication No. US20030109434A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Gordon, Brian  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER  
; FILE REFERENCE: 210121.572  
; CURRENT APPLICATION NUMBER: US/10/102,524  
; CURRENT FILING DATE: 2002-03-19  
; NUMBER OF SEQ ID NOS: 1863  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 645  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-102-524-645

Query Match 21.6%; Score 182.2; DB 14; Length 510;  
Best Local Similarity 69.5%; Pred. No. 3.9e-34;  
Matches 248; Conservative 0; Mismatches 104; Indels 5; Gaps 2;

Oy 492 CNGTGGGCGTNGTNTCTTCTTAATTCCTCCCAAGAAANNCTTAANGAANCCTCN 551  
Db 1 CTGTGGGATGTGTCTTCTTAATTCCTCCCAAGAAANNCTTAATGAATCTTCA 60  
Oy 552 CCAAAATACCTTAATATCTTGTGTAACCAAAACAACTTTTNGTTTACTANTC 611  
Db 61 GCAAAATATCTTAAATATCTTGTGTAACCAAAACAACTTTTNGTTTACTANTC 120  
Oy 612 CTGTGGGATTTAAGGGTCCCAATTTATTCGAAACCAATTTTCCCAACATANTT 671  
Db 121 CTGTGGATTTACGTCTTCTTAATTTATTCGAAACCAATTTTACCCACATANTT 180

QY 672 ACCATTTTACCTGTGTAAGCNCAGTNGTTTGCAANTCCGCAANAGTANTTTCC--C 729  
 DB 181 ACCATATTATCTTTGTTAATGACAGTGTATGCAATTCGCAAGAGATGATATCCATC 240  
 QY 730 CNGGNCCTTCCCGCCGACCTTGGGAAAAACGGGATNGTCCCGCCCTT---AAAAACA 786  
 DB 241 AGGCTCTATTCACCCGACAGATAGAAAAACAGAGATGCTACACCCATTAGAAAAACAA 300  
 QY 787 ACCTTCCCGCCCTTGGGCGCAAGNTTNTCCGCTTAATCCGAACATATAAAG 843  
 DB 301 TCTATCCACGCTATTTGGCAATGTGTAGTTCACTTAATTCGACATATAAATG 357

## RESULT 11

US-09-960-352-5962  
 ; Sequence 5962, Application US/09960352  
 ; Patent No. US20020137139A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 5962  
 ; LENGTH: 397  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 26-LIB188-014-Q1-E1-G5  
 US-09-960-352-5962

Query Match 10.0%; Score 84.6; DB 10; Length 397;  
 Best Local Similarity 54.1%; Pred. No. 1.4e-10;  
 Matches 211; Conservative 0; Mismatches 167; Indels 12; Gaps 4;

QY 438 ANCTATNGAAANTGATTANNAGCTGAATGCGCAACCTTANTTNCNGCCACGCTGG 497  
 DB 18 ATCAGTGAAGAAATATTAAGTCAATTAATGACAAACATTAATTAAGTCAAGTAATA 77  
 QY 498 GCTNGTNTCTTACTTANTCCCGCCAGAAANNCTTAANCAGANCTCCNCCAAA 557  
 DB 78 ATGAATTCCTGTTCTTCAATTCCTCCAGAGAAACT-----TGAATTCGCGAGAA 129  
 QY 558 TAACCTTAATATCTGTTGTAACCAAAACAAACCTTTTGTGTTAGTANTCCCTGGG 617  
 DB 130 TAATCTTCAATATACATTAATTAAGCAAAATGAGAGTTCT--GTTTACATCT--TTTGT 186  
 QY 618 ATTAAACGGGTCGCCAATTNATCCGAAACCCANTTTTCCCGCAACCATANTTACAT 677  
 DB 187 ATTGTCATTTCTAATCTTATTTAACTCAATTTTACCCCAACCATGATTAACATA 246  
 QY 678 TTAACCTGTGTAAGCAGTNGTTGCAATNCCGCAANCAAGTANTTTCC--CNGGCGC 736  
 DB 247 TTAACCTGTGTAAGCAGTNGTTGCAACTCAGCAACAGATGTAACATCAGGTTTC 306  
 QY 737 TTTCCCGCCGACCTTGGGAAAAACGGGATNGTCCCGCCCTTAATAAACACCTTCCCGC 796  
 DB 307 TATTAACCCAGACCTTAAGAAAAACAGATTAATTAACACCCATTAAACAAATCTCAT 366  
 QY 797 NCCTTGGCCAGGNTTNTTCCGCTCTAA 826  
 DB 367 GCTATTTCCCAATGATGAGTTTCAATCTAA 396

RESULT 12  
 US-09-960-352-13260/C  
 ; Sequence 13260, Application US/09960352  
 ; Patent No. US20020137139A1

; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 13260  
 ; LENGTH: 449  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 57-BOVMS1-016-Q1-E1-G2  
 US-09-960-352-13260

Query Match 9.3%; Score 78.2; DB 10; Length 449;  
 Best Local Similarity 56.5%; Pred. No. 5.2e-09;  
 Matches 139; Conservative 0; Mismatches 106; Indels 1; Gaps 1;

QY 597 TTNGTTTACNTANTCTTGGGATTTACGGGTCCCAATTNATCCGCAACCAATTTTC 656  
 DB 430 TTCTGTACCAATCTTTTGTATTTGTTATTTCTAATCTTAATAACCAATTTTA 371  
 QY 657 CCCCAACCATANTTACATTTTACCTGTGTAAGCNCAGTNGTTGCAANTNCCGCAAN 716  
 DB 370 CCCGACCATGATTAATCATATTAATGTGTAAGCAGATGTTGCACTCAGCAAG 311  
 QY 717 CAGTANTTTTCC--CNGGNCCTTCCCGCCGACCTTGGGAAAAACGGGATNGTCCCGCC 775  
 DB 310 CAGTGTAAACCATCAGTGTATTTATTTGTTATTTCTAATCTTAATAACCAATTTTA 251  
 QY 776 CTTAAAAACAACCTTCCCGCCGCTTGGGCGCAGGNTTNTTCCGCTTAATCCGAACA 835  
 DB 250 CCATTTAAAAACAACCTCATGTATTTCCCAATGTATATGTTCAATCAATCAATCTGACA 191  
 QY 836 ATAAAA 841  
 DB 190 ATGAAA 185

## RESULT 13

US-09-960-352-13016/C  
 ; Sequence 13016, Application US/09960352  
 ; Patent No. US20020137139A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 13016  
 ; LENGTH: 397  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 56-BOVMS1-016-Q1-E1-F8  
 US-09-960-352-13016

Query Match 8.3%; Score 69.6; DB 10; Length 397;  
 Best Local Similarity 56.2%; Pred. No. 5.8e-07;  
 Matches 163; Conservative 0; Mismatches 123; Indels 4; Gaps 3;

QY 553 CAATAATACCTTAATATCTTGTGTAACCAANCAAAACCTTTTNGTTTACATNTTCC 612  
 DB 397 CAGATAAGCTTCAATATATATATTAAGCAAAATGAGAGCTTCT--GTTTACATTAATCA 340



```

Oy 613 TTGGATTAAAGGGTCCCAATTNACCGAACCANTTTTCCCNACCAANTTA 672
Db 339 TT-GTATTCGTCTATTATTAACACTACTCTTAAACCTCACTTTACCCCAACGATGA 281
Oy 673 CCATTTTACCTTGGTAAGGCGNCAGTNGTTTGCANTNCGCAANAACAGTANTTTC-CCN 731
Db 280 CCAATTATTAACGTGTAAATGACAGATGTTTACCACTACGCAATGCTGTAGTAACATCA 221
Oy 732 GGCNCTTTCCCCCGANCTTGGGAAAAACGGGATNGTCCCTCCCTTAAAAACAACCTT 791
Db 220 GGTATATATTCACCCGCAATTGGATAAAAACAGACTGTGATCACCACATTTAAAAACAAC 161
Oy 792 CCCCCNCTTTGGCCGAGNNNTNTTCCCGCTAAATCCGAACAATAAA 841
Db 160 CTCATGCTATTTCCCAATCTAATAGTTTCAATCTAATCTTGAACAATGAAA 111

RESULT 14
US-10-311-455-914/C
Sequence 914, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Cytosine Methylation
FILE REFERENCE: 5013.1014
CURRENT FILING DATE: 2002-12-16
PRIORITY FILING DATE: 2001-07-02
PRIORITY FILING DATE: 2000-06-30
PRIORITY FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 914
LENGTH: 18133
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-914

Query Match 7.3%; Score 61.8; DB 12; Length 18133;
Best Local Similarity 40.1%; Pred. No. 0.0036;
Matches 174; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

Oy 5 AAAAAAAAAATTATTAAATTTCTCTATTANCMTCCTCCGCAANACATTTATNACCCTATN 64
Db 16239 AAAAAACAATACATACCTACCATTAATAAACGTCGCAAAATTAATATATTTTATACATTTT 16180
Oy 65 NCNCNCNANTTTNNAANAANTACCTTNNNTTAAAAAACCTNGGAAAAAATAATNGC 124
Db 16179 CACCTCTAACAATAATCATTTTCAATATATATCATATATATTAATTAATAATATATTTT 16120
Oy 125 AAATATNTTAACTTCTGTAAGGAAAGAAATTTTATCCAGANGACGAAANCTTNTAATT 184
Db 16119 CTTACTTTTACTCTAATTAATAAAAAACAATATTTTAAAAAATACTTTTCCACTTTAATA 16060
Oy 185 NGANNTNAAATTAATNTNGAANCGGNCNGAAGCAACCAANTCTNATGTCGAATTAATCC 244
Db 16059 TAAATATAAAAATTCATTAACCATTTTCATATATTAACCAAAATTTTCACTTACCAAAAA 16000
Oy 245 TNAANGAGGNNTTNANNACTAATNCCNGATTTTCCAATANGAANCCNNNTTAA 304
Db 15999 AATCTTACATCTTACAAATTAACAATATTTACTATTTTCAATTTTAAATACAAATAAAA 15940
Oy 305 NTNTTTTAAATTTTAAATAACCCNGTTCGAACCCCGAATCAATTCCTTNTAATTTGA 364
Db 15939 TCCAAATTAACACAAATAAACAATAAATAAATAAACAATACTATTTCTTTATTTTACA 15880

```

```

OY      365 TGGGGGAAAAAAAAAATNCNGTCCNNATACCGNNGAAGNCAAAATTTTAAATTTTAAACC 424
Db      15879 ATCATCTTACTATCATTTATTTATTTCAATTCACATTAATAACAAAATATCTAATAC 15820

OY      425 CCCTAATTTTAAAA 438
      | | | | |
Db      15819 TACAAATTTTAAAA 15806

RESULT 15
US-10-311-455-1922/c
; Sequence 1922, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Cytosine Methylation of Cytosine Methylation Sites
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1922
; LENGTH: 18218
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1922

```

```

Query Match          6.1%; Score 51.8; DB 12; Length 18218;
Best Local Similarity 42.0%; Pident No. 0.094;
Matches 116; Conservative 0; Mismatches 160; Indels 0; Gaps 0

QY 44 AAANCATTAATTNNCCCTATNNCCNCGAATTTTANANAANTACCTTTNNTTTAAAAA 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4063 AAAAAATTATTTTAAAAAACTTTAAAAATACCTTAACCAAAAAATTTATATTAATAAA 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 104 CCTNGGAAAAAAAATAATNGCAAAATANTTAACTTCTTGAAAAANGAAATTTNACCA 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4003 CAATATAAAATAAAAAAACTAAAAAATTTAAAAATATATTTTAAAAATAAAAACATTA 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 164 NGGACNGAANCNTTNTTAAATTTNGAANTTAAATTTATNTNGAANAANGCGCNCNGAA 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3943 ATTTCTTAACATAATATATATTTAAAAACATAAAATATTTTAAAAATATATACCTTT 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 224 ANCTTNAATGTCATTAATCCTNANGAGGNNTTTANNACTAATNCCNGATTTTCCA 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3883 ATAAATACATCAATTTTATATAAAAAATATCATATTTTAAATCTCAATTTTATTA 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 284 ATANGGAANCCCNNTTAAAAAATNTTTTNAATTTTAA 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3823 TTTTATAATTCCTATATTAACCTTTTTTCTTAA 3788
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: November 27, 2003, 12:53:10
Job time : 320.882 secs

```

**THIS PAGE BLANK (USPTO)**

Qy	Db	Qy	Db
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

```

Db      61 TATTNNCCNCGANTTTTNAAAAATPACCTTTNNNTTAAAAAACCCTNGAAAAAATAA 120
Qy      121 TNGCAATANTTAACTTTCCTTGAAAANGAAATTTNTACCAANGACGAAANCNTNT 180
Db      121 TNGCAATANTTAACTTTCCTTGAAAANGAAATTTNTACCAANGACGAAANCNTNT 180
Qy      181 AATTGAAATNTAAATTAATNTTNGAANGGCGNCGAAACCAANTTAAGTCCAATT 240
Db      181 AATTGAAATNTAAATTAATNTTNGAANGGCGNCGAAACCAANTTAAGTCCAATT 240
Qy      241 ATCTTNAAGAGGANNTTTNNANNACTAATNCCNGATTTTCCAAATNGAANCCNNNT 300
Db      241 ATCTTNAAGAGGANNTTTNNANNACTAATNCCNGATTTTCCAAATNGAANCCNNNT 300
Qy      301 AAAANTNTTTTAAATTTTAAAAAATAACCCNGTNTCCAAACCCNGATCANATTCCTTNAAT 360
Db      301 AAAANTNTTTTAAATTTTAAAAAATAACCCNGTNTCCAAACCCNGATCANATTCCTTNAAT 360
Qy      361 TGGATTGGGGAAAAAATNCGTTCNNATACCGNAAANGCAANTTTTAAATTTTAA 420
Db      361 TGGATTGGGGAAAAAATNCGTTCNNATACCGNAAANGCAANTTTTAAATTTTAA 420
Qy      421 ACCCCCTANTTTTAAAAAANTNGAATTAANNAGCTGAATGGCAACCCCTAN 480
Db      421 ACCCCCTANTTTTAAAAAANTNGAATTAANNAGCTGAATGGCAACCCCTAN 480
Qy      481 TTNCNGGCCACGNGTGGGCTNTGTTCTTAATTATTCCTTAAATTCCTTAA 540
Db      481 TTNCNGGCCACGNGTGGGCTNTGTTCTTAATTATTCCTTAAATTCCTTAA 540
Qy      541 CNGAANTCCNCCAAATTAACCTTAATATCTTGTGTAACCAANCAAACTTTTNG 600
Db      541 CNGAANTCCNCCAAATTAACCTTAATATCTTGTGTAACCAANCAAACTTTTNG 600
Qy      601 TTTACANTATCTTGGGATTTAAAGGGTCCCAATTTNATCCNGAACCCANTTTTCC 660
Db      601 TTTACANTATCTTGGGATTTAAAGGGTCCCAATTTNATCCNGAACCCANTTTTCC 660
Qy      661 NAACATANTTAACATTTTACCTTGTGTAAGGCGNCACTGTTGCAATNCCGCAANCACT 720
Db      661 NAACATANTTAACATTTTACCTTGTGTAAGGCGNCACTGTTGCAATNCCGCAANCACT 720
Qy      721 AATNTTCCCGNCGCTTTCCCGGCAACCTTGGGAAAAAGGATNTGTCCTTAA 780
Db      721 AATNTTCCCGNCGCTTTCCCGGCAACCTTGGGAAAAAGGATNTGTCCTTAA 780
Qy      781 AAAACACCTTCCCGNCGCTTGGCCGCAAGNNTTNTCCGCTAATCCGAACATATA 840
Db      781 AAAACACCTTCCCGNCGCTTGGCCGCAAGNNTTNTCCGCTAATCCGAACATATA 840
Qy      841 AAG 843
Db      841 AAG 843

```

RESULT 2  
US-08-896-164-19  
; Sequence 19, Application US/08896164  
; Patent No. 6218521

GENERAL INFORMATION:  
; APPLICANT: OARITA, Yuichi  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED  
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR  
; TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER  
; NUMBER OF SEQUENCES: 87  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/896.164
; FILING DATE: July 17, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6218521man D. Hanson
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5499 - JBL/NDH/SLH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-896-164-19

Query Match      51.7%; Score 436.2; DB 3; Length 714;
Best Local Similarity 73.1%; Pred. No. 1.2e-108;
Matches 516; Conservative 0; Mismatches 187; Indels 3; Gaps 3;

Qy      2 GCCAAAAAATTTTAAATTTTCTATTAAATTCCTCCNCAAAACATTAATTNACCT 61
Db      2 GCCAAAAAAGTTATTTTAAATTTTCTATTAAATTCCTCCAAAGCATTTATCTCT 61
Qy      62 ATNNCCNCGAANTTTNAAAANTACCTTNNNTTAAAAAAACCTNGAAAAAATAT 121
Db      62 ATATCTCACTAATTTTAAAGAAATACATTTAGTATTGAAAAAAGTAAAGATATA 121
Qy      122 NGCAATANTTAACCTTCTGAAAANGAAATTTNTACCAANGACGAAANCNTNTA 181
Db      122 TGCAGATATTAATTAACATTAATGAAAAAGAAATTTATACAAAGAGCTGAGAGTATA 181
Qy      182 ATTNCAANTNAAATTAATNTTNGAANGGCGNCGAAGCAACCAANTTATGCTCAATTA 241
Db      182 AATTGAATGAGATTATATTTGAAAACTGATCGAAGCAAACTTATTTGTTCAATTA 241
Qy      242 TCTTNAANGAGGANNTTNANNACTAATNCCNGATTTTCCAAATNGAANCCNNNTTA 301
Db      242 TNCCTTAATGAGTGGTGTATGACTAATACATGATTTTCAAGAGGAAACCATGTTA 301
Qy      302 AAANTNTTTTAAATTTTAAAAATACCCNGTNTCCAAACCCNGATCANATTCCTTNAATTT 361
Db      302 AAAATATTTTAAATTTTAAAAATAGCTGTGTCAAGCTGTGATCATATTTCTTTATTT 361
Qy      362 GGATTTGGGAAAAAATNCGTTCNNATACCCNNGAANGCAANTTTTAAATTTTAA 421
Db      362 TGATTTGGGAAAAAATACGTGTCTGTGATACGAAATGCAAAATTTTGAATTTTAA 421
Qy      422 CCCCCCTAATTTTAAANCTATNGAANAANTNGATTANNAGCTGATTC-CAACCTAN 480
Db      422 TCTCACTAATTTTAAANCTAATGAAATGATTAATGATGAAGTCAACACTTAA 481
Qy      481 TTNCNGGCCACGNGGCGTNTGTTCTTACTTANTCCCGCAA-GGAANNCCCTTAA 539
Db      481 TTACGGCCACCTGTGGCATGTGTTCTTACTAGTTCGCCAAGGAAAACTCTTAA 541
Qy      540 NCGAANTCCNCCAAATTAACCTTAATATCTTGTGTAACCAANCAAAACCTTTTNT 599
Db      542 AATTGAATTTGAGGAGATATATCTTAATATCTTGTGTAACCAAAAGCTTTTNT 601
Qy      600 GTTACATNTA-NTCTTGGGATTTAAAGGGTCCCAATTTNATCCNCAACCAATTTTCC 658
Db      602 GTTACATAGTTCTTGGGATTTTACGTGTTCTTATTTATTTGAAACCAATTTTACC 661
Qy      659 CNAACATANTTAACATTTTACCTTGTGTAAGGCGNCACTGCTTGC 704

```

Db 662 CCAGACCAATTAACCATTAATCTTGTGTGACAGTGTG 707

## RESULT 3

US-08-896-164-20  
; Sequence 20, Application US/08896164  
; Patent No. 6218521  
; GENERAL INFORMATION:  
; APPLICANT: OSAITA, Yuichi  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED  
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR  
; TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER  
; NUMBER OF SEQUENCES: 87  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felte & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/896,164  
; FILING DATE: July 17, 1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6218521man D. Hanson  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 687 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-896-164-20

Query Match 48.8%; Score 411.8; DB 3; Length 687;  
Best Local Similarity 71.8%; Pred. No. 4,4e-102;  
Matches 495; Conservative 0; Mismatches 190; Indels 4; Gaps 3;

Qy 15 ATTNAATTCTTATTAAACNTCCGNCNCAANCAATTATTNAACCTATNCCNCGNAN 74  
Db 1 ATTTTAAATTTCTTAAACATCTTCTCAAGCATTTATTTATCTTATATCTCACTGA 60  
Qy 75 TTTNANAANTACCTTTNNNTTTAAAAAAGCTNGGAAAAAATAATGCAATANTTA 134  
Db 61 TTTTAAATAATACCTTATGATTTAGAAAACTAGAAAAAAGATTAATGCAATATTA 120  
Qy 135 CCTTCTGAAAAAGAAATTTNTACCAANGACNCAANCTTTAATTGAAATNAAA 194  
Db 121 ACTTACATGAAAAAGAAATTTATTAACAAGACGTGAGAACTTTAATGAAATGGA 180  
Qy 195 TTATANTTNGAAAGCGNCGNCAACCAACTTAATGCTCAATTTCTTAANGAGGG 254  
Db 101 TTATATATTGAAACATGATCTGAAAGCAAACTTTATGTTCAATTTATTTAATGATGG 240  
Qy 255 NNTTANNAATTAATNCCNGATTTTCCATANGAANCCNNTTAAANNTTTTAT 314  
Db 241 TGTTTTATGACTTAATACACTGATTTTTCATTAAGAAACCAATGTTAAATAATTTTAT 300  
Qy 315 TTTAAATAATACCCNGTNTCAACCCCGATCANATTCCTTNAATTGGATTGGGAAA 374  
Db 301 TTTAAATAATACCCGTGTGTTCAAGCTGTGATCATATTTCTTTAATTTGATTTGGGAGA 360

Qy 375 AATNCGTTCNNATACNNGAANNGCAANTTTTAAATTTTAACTCCCTANTTTT 434  
Db 361 AATACGTCTTCTGATGACAGAAATGCAAAATTTTGAATTTTAACTGCTAATTTT 420  
Qy 435 AAAANCTATNGAAAAANTGATTAANGACTGAATTC-CAACCTTANTTNGGCCACCN 493  
Db 421 AAGAACTATGAGAAATGATTAATGACATGAAAGTGAACAACACTAATTTACTGGCAGCT 480  
Qy 494 GTGGCGTNGTNTTCTTACTTANTCCCCCAAGAAANNCCTTAANGAANCTCCNC 553  
Db 481 GTTGGCATGTGTTCTTACTAGTTCGCCAAGAAACCTTTAACTGAATCTTCAGC 540  
Qy 554 AAATTAACCTTAATATATCTTGGTGAACAAANCAAACTTTTNGTTTACNTANTCCT 613  
Db 541 NGATTAACCTTAATATATCTTGTGACCAAC--AAACTTTTGTGTTTACATGTTCT 598  
Qy 614 TGGATTTTAACGGGTCCCAATTTNATCNGAACCANTTTTCCCNAACTANTTAC 673  
Db 599 TTGATTTTACGTGTTCTTAATTTTATCTGAAACCTCATTTTCCAGACATAATTAAC 658  
Qy 674 CAT-TTTACTTGTGTAAGGCNCACTGTT 701  
Db 659 CCTATTTAACTTTGTTATGACAGTGTGT 687

## RESULT 4

US-09-313-294A-1553/C  
; Sequence 1553, Application US/09313294A  
; Patent No. 6476212  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath V.  
; APPLICANT: Ito, Laura Y.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
; FILE REFERENCE: PL-0017 US  
; CURRENT APPLICATION NUMBER: US/09/313,294A  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 7600  
; SOFTWARE: PERL Program  
; SEQ ID NO 1553  
; LENGTH: 173  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. 6476212 700551132H1  
US-09-313-294A-1553

Query Match 13.0%; Score 110; DB 4; Length 173;  
Best Local Similarity 71.3%; Pred. No. 7.3e-21;  
Matches 122; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 8 AAAATTTATTTAAATTTCTTAAACNTCCGNCNCAANCAATTATTNACCTATNCCN 67  
Db 172 AAAAGATTTATTTCTATTAACATCTTCTGCAAGCATTTTATCTATCT 113  
Qy 68 CNGCNGATTTNANAANTACCTTTNNNTTTAAAAAAGCTNGGAAAAAATAATGCAAA 127  
Db 112 CACTGAATTTTAAAGAAATPAACATTAATGTAATGAAAAAATAGATTAATGACGA 53  
Qy 128 TANTTAACCTTCTGAAAGAAATTTNTACCAANGAANGAANCTT 178  
Db 52 TAAATTAACCTTAACATGAAAAAAGAAATTAATPAACAAAGACTGAGAACGTT 2

## RESULT 5

US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.

```

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT29pc-F15
; US-08-232-463-14

Query Match
Best Local Similarity 2.1%; Score 48.8; DB 1; Length 7218;
Matches 8; Conservative 201; Mismatches 174; Indels 0; Gaps 0;

QY 456 TANNAGCTGAATGGCAACCTTANTNCGCCACGCGTGGCAGTGTTCCTTACTT 515
DB 1060 TTGCGATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1119
QY 516 ATCCGCCCAAGGAANNCTTANCGAANCTCCNCGCAAAATTAACCTTANATCTT 575
DB 1120 YTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1179
QY 576 GGTAAACCAANCAAAACCTTTTNGTTTACNTANTCTTGGATTAAACGGCTCCCAT 635
DB 1180 YTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1239
QY 636 TTNATCCNAGCCATTTTCCCCNAGCCATTTTACATTTTACCTTGTAGGCGNCA 695
DB 1240 YTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1299
QY 696 GTNGTTGATNCGCAANCAAGTANTTTCCCGGCGNCTTTCCCGGAGCTTGGGA 755
DB 1300 YTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1359
QY 756 AAAAGGATNGTCCCGCTTAAACCACTCCCGGCTTGGCCGAGGNTTN 815
DB 1360 YTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1419
QY 816 TTCCGCTTAATTCGAGACATA 838
DB 1420 YTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1442

```

```

RESULT 6
US-09-302-812-38
; Sequence 38, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARC) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; FILE REFERENCE: NIND 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; EARLIER FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO: 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; US-09-302-812-38

Query Match
Best Local Similarity 5.3%; Score 44.6; DB 4; Length 29793;
Matches 145; Conservative 0; Mismatches 206; Indels 4; Gaps 1;

QY 5 AAAAAAATTTTAAATTTTCTTATTAATNCTTCGCGCAANCAATTTTACCCATN 64
DB 18899 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 18958
QY 65 NCCNCCGANTTTNANAATTAACCTTTNNTNTTAAAAAACCCTGAAAAAAT---AA 120
DB 18959 ACTGAAAAATTAAGCTTATTAAGCTTATTAAGCTTATTAAGCTTATTAAGCTT 19018
QY 121 TNGCAATTAATTAACCTTCTTGAAGAAATTTTACCAAGACGAAACCTTNT 180
DB 19019 TTAATTAATTAACCTTCTTGAAGAAATTTTACCAAGACGAAACCTTNT 19078
QY 181 AATNGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
DB 19079 ACGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19138
QY 241 ATCCNANAGAGGNTTTNANNACTAATNCCNGATTTTCAATANGAANCCNNNT 300
DB 19139 TTTCTTTTGTACTTCTTAATAAATTAATTAATTAATTAATTAATTAATTAAT 19198
QY 301 AAAATNTTTTATTTTAAATTAACCGGNTTCAACCCCGGATCAATTCCTT 355
DB 19199 TGAACATGTTTGAACAAAAAATTTTGAACAAACCCCAAAATTAATTCCTT 19253

RESULT 7
US-09-511-477-38
; Sequence 38, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARC) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; FILE REFERENCE: NIND 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; EARLIER FILING DATE: 2000-02-23
; EARLIER APPLICATION NUMBER: 09/302,812
; NUMBER OF SEQ ID NOS: 38

```

SEQ ID NO 38  
LENGTH: 29793  
TYPE: DNA  
ORGANISM: Caenorhabditis elegans  
FEATURE:  
US-09-511-477-38

Query Match 5.3%; Score 44.6; DB 4; Length 29793;  
Best Local Similarity 40.8%; Pred. No. 0.025; Mismatches 206; Indels 4; Gaps 1;  
Matches 145; Conservative 0; Mismatches 206; Indels 4; Gaps 1;

QY 5 AAAAAAATTTATTTAATTTCTATTAACNTCTCCNCAANCAATTATTNACCTATN 64  
DB 18899 AATATTGATTTTATGAAAAAATAATACGTTTGACCTTAAAAAATATATGTTAA 18958  
QY 65 NCNCNCNGANTTTNANAAATACCTTTNTNTTAAAAAACCCTNGAAAAAAT---AA 120  
DB 18959 ACTTGAAAAATACGTTTATGCTTATATTAATAATCGGTTTCAATTAATTTAA 19018  
QY 121 TNGCAATANTTAACCTTCTGAAANGAATTTNTCCANGACNGAANCTNTT 180  
DB 19019 TTTAAATATCTACCTTTGGTCAAAATCAGACATTTTGAACATGACATGTTAT 19078  
QY 181 AATTGAAATTAATTAATTAATTTGAAANGCGCNCNGAACAACCAACTTATGTCCAAT 240  
DB 19079 ACGTTGAATATACCTATGTTGAAAAATGAAAAATTTGAAGACAGTGAATTTAGTTT 19138  
QY 241 ATCTTGAANGAGGANTTTNANNACTAATNCCNGATTTTCCAAATPANGAANCCNNNTT 300  
DB 19139 TTTCTTTTCTGCTACTCTTAAAAAATACCTTCAATTAATTTTGAAGACTAATTTT 19198  
QY 301 AAAATNTTTTATTTTAAATTAACCCNGTTCACACCCNGATCANATTCCTT 355  
DB 19199 TGAACATGTTTGAACAAAAAAGATTTGAAAAACCCCAAAAAAATCTACTTT 19253

RESULT 8  
US-09-511-507-38  
Sequence 38, Application US/09511507  
Patent No. 6395343  
GENERAL INFORMATION:  
APPLICANT: JACOBSON, Myron K.  
APPLICANT: JACOBSON, Elaine L.  
APPLICANT: AME, Jean-Christophe  
APPLICANT: LIN, Winston  
TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PAR) EN  
TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
FILE REFERENCE: NIND 201  
CURRENT APPLICATION NUMBER: US/09/511,507  
CURRENT FILING DATE: 2000-02-23  
PRIOR APPLICATION NUMBER: 09/302,812  
PRIOR FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 38  
SEQ ID NO 38  
LENGTH: 29793  
TYPE: DNA  
ORGANISM: Caenorhabditis elegans  
FEATURE:  
US-09-511-507-38

Query Match 5.3%; Score 44.6; DB 4; Length 29793;  
Best Local Similarity 40.8%; Pred. No. 0.025; Mismatches 206; Indels 4; Gaps 1;  
Matches 145; Conservative 0; Mismatches 206; Indels 4; Gaps 1;  
QY 5 AAAAAAATTTATTTAATTTCTATTAACNTCTCCNCAANCAATTATTNACCTATN 64  
DB 18899 AATATTGATTTTATGAAAAAATAATACGTTTGACCTTAAAAAATATATGTTAA 18958  
QY 65 NCNCNCNGANTTTNANAAATACCTTTNTNTTAAAAAACCCTNGAAAAAAT---AA 120  
DB 18959 ACTTGAAAAATACGTTTATGCTTATATTAATAATCGGTTTCAATTAATTTAA 19018

QY 121 TNGCAATANTTAACCTTCTGAAANGAATTTNTACCAANGACNGAANCTNTT 180  
DB 19019 TTTAAATATCTACCTTTGGTCAAAATCAGACATTTTGAACATGACATGTTAT 19078  
QY 181 AATTGAAATTAATTAATTAATTTGAAANGCGCNCNGAACAACCAACTTATGTCCAAT 240  
DB 19079 ACGTTGAATATACCTATGTTGAAAAATGAAAAATTTGAAGACAGTGAATTTAGTTT 19138  
QY 241 ATCTTGAANGAGGANTTTNANNACTAATNCCNGATTTTCCAAATPANGAANCCNNNTT 300  
DB 19139 TTTCTTTTCTGCTACTCTTAAAAAATACCTTCAATTAATTTTGAAGACTAATTTT 19198  
QY 301 AAAATNTTTTATTTTAAATTAACCCNGTTCACACCCNGATCANATTCCTT 355  
DB 19199 TGAACATGTTTGAACAAAAAAGATTTGAAAAACCCCAAAAAAATCTACTTT 19253  
RESULT 9  
US-08-916-421B-1/c  
Sequence 1, Application US/08916421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Bull et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ  
Patent No. 6503729  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA  
ORGANISM: Methanococcus jannaschii  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98343)..(98343)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (10398)..(10398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (148948)..(148948)

```
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385) ..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989) ..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995) ..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980) ..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220) ..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814) ..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398) ..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418) ..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837) ..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993) ..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226) ..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167) ..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241) ..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992) ..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708) ..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081) ..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203) ..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435) ..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442) ..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652) ..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684) ..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455) ..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676) ..(779676)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc feature
LOCATION: (855539) ..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619) ..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830) ..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846) ..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1115881) ..(1115881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881) ..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988) ..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224) ..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1345473) ..(1345473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491) ..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091) ..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020) ..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912) ..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734) ..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998) ..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854) ..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

Query Match 4.9%; Score 41.6; DB 4; Length 1664976;  
Best Local Similarity 42.6%; Pred. No. 0.79;  
Matches 121; Conservative 2; Mismatches 160; Indels 1; Gaps 1;

```
QY 96 TTAATAAACCTGNGGAAAAAATATATNGCAATANTTTACCTTCTGTAAGAAAGCAATT 155
DB 1150605 TTAATAAGACTGCTGACGCAATAGTGAATGGACACCAAGAAATAGAAACC 1150546
QY 156 TTTACCAANGACNGAANCAANTTTAATTNGAANTTAATTAATTNGAANGCGCNC 215
DB 1150545 TTYACAAATCTTAAGTAAGAACTTCAAAAAATTAACGTTCAATGATATATAT 1150486
QY 216 NGAAACCAANTCTNVTGTCCTCAATTATCTTAANGAGGNNTTT-NANNACTAATNCCN 274
DB 1150485 GGACAACTAATTTCAATCCGTATATCTGTAGATTACGTTAGRAAACTCGAACAT 1150426
QY 275 GATTTCCAAATANGAAGAAACCCNNNTTAANTTTTAATTAATAACCCGNTATC 334
DB 1150425 CAATTCAAATPAAGAAAGACGACGTTAAATTTGTTGAAGATTAATCAATTAAT 1150366
QY 335 CAACCCGATCATNATTCCTTNAATTTGGATTGGGAAAAAAT 378
DB 1150365 TAAATTTAATTAATTAATTAATGTAATGTAATGTAATTAATTAATTAATTAAT 1150322
```





```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match
Best Local Similarity 4.7%; Score 39.4; DB 1; Length 5852;
Matches 85; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 16 TTNAATTCCTATTGACGCTCCGCAAAACATTATTNACCCATNCCNCCGANT 75
DB 5751 TTGATATTTTATATATGCTATTTGTTGTTTCTTACTATATTTCTATTTT 5692
QY 76 TTNANAANTACCTTNNNTTAAAAAACCCTNGAAAAAATAATNGCAATATTAA 135
DB 5691 TTTATTAATTAATTAATTAATTTTAAATTAATAAATAAATAAATAAATTT 5632
QY 136 CTNCTTGAAANGAAATTTTACCAANGACNGAAACNTTNTAATNGAANTAA 195
DB 5631 ATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5572
QY 196 TATANTTNGAANC 211
DB 5571 TAAAAAATGTAAAGG 5556

RESULT 13
US-08-446-855A-1/C
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: Nixon & Vanderhye PC
STREET: 1100 No. 5849573ch Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
```

```
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match
Best Local Similarity 4.6%; Score 39; DB 2; Length 8920;
Matches 87; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 6 AAAAAATTTATTTTATTCCTATTGACGCTCCGCAAAACATTATTNACCTATNN 65
DB 8732 AATAATTTATTTAATAATTAATAATTTGTAATGAAAAACCATTTGGTTATAC 8673
QY 66 CACNCGANTTTNANAANTACCTTNNNTTAAAAAACCCTNGAAAAAATAATNGCA 125
DB 8672 ATATGAGTAATTAATAATTTGTAATATTAACAAAAATTATTAATAATCAAT 8613
QY 126 AATANTTACCTTCTGAAANGAAATTTTACCAANGACNGAAACNTTNTAATTN 185
DB 8612 ATTAATTAATGATCATATTTAATAAACAATTTATTAATTTAAACGTAATATTA 8553
QY 186 GAANTTAATTAATNTNGAAA 207
DB 8552 AATAAATTAATATATATATATA 8531

RESULT 14
US-09-150-741-1/C
Sequence 1, Application US/09150741
Patent No. 6183996
GENERAL INFORMATION:
APPLICANT: Stewart et al.
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
Patent No. 6183996
TITLE OF INVENTION: Synthetase II
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/150,741
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PL6380
EARLIER FILING DATE: 1992-12-16
EARLIER APPLICATION NUMBER: AU93/00617
EARLIER FILING DATE: 1993-12-02
EARLIER APPLICATION NUMBER: 08/446,855
EARLIER FILING DATE: 1995-07-06
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8920
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match
Best Local Similarity 4.6%; Score 39; DB 3; Length 8920;
Matches 87; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 6 AAAAAATTTATTTTATTCCTATTGACGCTCCGCAAAACATTATTNACCTATNN 65
DB 8732 AATAATTTATTTAATAATTAATAATTTGTAATGAAAAACCATTTGGTTATAC 8673
QY 66 CACNCGANTTTNANAANTACCTTNNNTTAAAAAACCCTNGAAAAAATAATNGCA 125
DB 8672 ATATGAGTAATTAATAATTTGTAATATTAACAAAAATTATTAATAATCAAT 8613
QY 126 AATANTTACCTTCTGAAANGAAATTTTACCAANGACNGAAACNTTNTAATTN 185
DB 8612 ATTAATTAATGATCATATTTAATAAACAATTTATTAATTTAAACGTAATATTA 8553
```

Qy 186 GAANTNAATTTATNTNGAAA 207  
 Db 8552 AATTAATTAATTAATTAATTA 8531

## RESULT 15

US-09-426-290-1  
 ; Sequence 1, Application US/09426290  
 ; Patent No. 6410712  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berglind Ran Olafsdottir  
 ; APPLICANT: Jeffrey Guicher  
 ; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE  
 ; FILE REFERENCE: 2345,2001-000  
 ; CURRENT APPLICATION NUMBER: US/09/426,290  
 ; NUMBER OF SEQ ID NOS: 24  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 168575  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (21181)...(21403)  
 ; NAME/KEY: CDS  
 ; LOCATION: (95252)...(95430)  
 ; NAME/KEY: CDS  
 ; LOCATION: (101753)...(101996)  
 ; NAME/KEY: CDS  
 ; LOCATION: (110324)...(110439)  
 ; NAME/KEY: CDS  
 ; LOCATION: (124058)...(124278)  
 ; NAME/KEY: CDS  
 ; LOCATION: (127009)...(127130)  
 ; NAME/KEY: CDS  
 ; LOCATION: (128910)...(129139)  
 US-09-426-290-1

Query Match 4.6%; Score 38.6; DB 4; Length 168575;  
 Best Local Similarity 50.0%; Pred. No. 2.1;  
 Matches 65; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 99 AAAAAGCTTNGAAAAAATATATNGCAATTAACCTTCTGTGAAAAAGAAATTTT 158  
 Db 125818 AAAAAGCTTGTGAAAAAAGAACTTGAAATGTTTCTGTACACACACATTTT 125877  
 Qy 159 ACCAAGACGACGAAGAACNTTNTAATTNGAANTNAATTAATNTATNTNGAAGCGCNCNGA 218  
 Db 125878 TCAAAATGCTTAATTAATTTATTCCTTGCAATGAATTTGTTTCTTCTCCACAAA 125937  
 Qy 219 AACCAACTT 228  
 Db 125938 AACCAATTCCT 125947

Search completed: November 27, 2003, 12:32:32  
 Job time : 65.2227 secs

**THIS PAGE BLANK (USPTO)**